Characterization of malaria sexual stage development in the human host

A dissertation presented by

Regina Joice

to

The Department of Immunology and Infectious Diseases

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in the subject of

Biological Sciences in Public Health

Harvard University

Cambridge, Massachusetts

April 2013

Dissertation Advisor: Dr. Matthias Marti Regina Joice

Characterization of malaria sexual stage development in the human host

ABSTRACT

Due to an increase in malaria control programs in the last decade, the world has witnessed dramatic reductions in the number of infections and deaths caused by the malaria parasite. With malaria eradication on the global health agenda, a shift toward transmission-focused research has led to a renewed focus on a previously neglected stage of malaria: the sexual stage (gametocyte). Malaria's sexual stages are the only stages in the human host that are transmitted to the mosquito vector, and are therefore of critical importance for blocking transmission of this devastating disease.

The process through which developing gametocytes sequester outside of the bloodstream during their 8-10 day maturation is not well understood and stands to be exploited as a potential target for therapeutic intervention. In Chapter 1, we discuss the current state of knowledge on the development of these stages in the human host. In Chapter 2, we investigate anatomical enrichment sites for developing gametocytes in the human host using autopsy tissue from cases of fatal malaria. Immunohistochemistry (IHC) and quantitative reverse transcriptase PCR-based assessments identified the bone marrow as a preferential enrichment site of developing gametocytes. Co-localization with host proteins revealed the enrichment of gametocytes inside the extravascular space of the bone marrow, often observed in contact with erythroblastic island structures. *In vitro* experiments with erythrocyte precursor cells, as well as *in vivo* co-

localization studies demonstrated that gametocytes *can* develop within the cells of the hematopoietic system of the bone marrow.

In Chapter 3, we present an assay and analysis tool for inferring the presence of young and mature asexual and sexual stages in the peripheral blood of infected patients based on gene expression data. We apply this assay to malaria patient cohorts and *in vitro* drug perturbation time course experiments, and demonstrate its use in identifying young and mature gametocyte carriers, as well as characterizing the effect of a given perturbation on parasite development. This body of work aims to contribute to the overall knowledge base for malaria's elusive gametocytes as well as to establish tools for performing future assessments on these transmissible stages.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Matthias Marti, for his support and encouragement throughout my time in his laboratory. I am grateful for the opportunities he has encouraged me to pursue, and the leadership roles he has encouraged me to take on. I thank him for consistently believing in my abilities, even when I've doubted them myself!

I would also like to thank several other faculty members who were highly influential in my growth as a scientist. For their encouragement and advising, I would like to thank Dr. Manoj Duraisingh and Dr. Dyann Wirth, who have helped guide my dissertation work from Day 1. I also thank Dr. Dyann Wirth for giving me the opportunity to participate in a variety of symposia focused on malaria eradication, as these experiences were very influential to my overall growth as a public health scientist. I would like to thank Dr. Danny Milner for his support of my project, his encouraging words along the way, and for making me laugh even at my most stressed moments. I would like to thank Dr. Terrie Taylor who served as an additional advisor for the field part of my project and has been tremendously supportive of my work. I would like to thank Dr. Curtis Huttenhower for his support and helpful advice along the way. I would also like to thank Dr. Marc Lipsitch for his mentorship and support throughout my time as a student, and for giving me a home within the Center for Communicable Disease Dynamics.

I would also like to thank my Dissertation Advisory Committee: Dr. Barbara Burleigh, Dr. Peter Weller and Dr. Marc Lipsitch, who were tremendously helpful in their suggestions throughout my time as a student, and have been very supportive of my work and my professional development. I would also like to thank the members of my Dissertation Exam Committee, for

participating in my defense: Dr. Barbara Burleigh, Dr. Caroline Buckee, Dr. James Mitchell, and especially Dr. Jim Kazura for traveling in from out of town.

I would like to thank the large group of individuals who are based out of the Malawi site, in particular Jimmy Vareta, Dr. Jacqui Montgomery and Dr. Karl Seydel. My projects could not have functioned or been as fun without them. I would also like to thank my labmates for their support along the way and for making the long days in lab enjoyable. I'd also like to thank the members of the other labs in the department, particularly the members of the Duraisingh, Wirth, Fortune and Burleigh labs, for providing advice, support, reagents and good break room lunch conversation. In particular, I would like to thank Dr. Sandra Nilsson, Vagheesh Narasimhan, Tom Burke and Evan Meyer, who directly aided in the work presented in this dissertation, and Dr. Pierre-Yves Mantel, Dr. Kacey Caradonna and Dr. Johan Ankarklev for providing helpful discussion. I would finally like to thank my family and friends who have helped me get through the tough times and celebrated with me during the joyous ones. In addition to my supportive parents Alan and Gerri and brother and sister Gregory and Veronica, I thank my friends Derrick Cordy, Rupal Shah, Amy Bei, Felisa Nobles, and Nadia Trowers for being particularly supportive and encouraging throughout my time as a PhD student.

DEDICATION

This dissertation is dedicated to my grandmother, Justina Williams, who in 1958 became the first African-American hired at North Carolina State University above the level of custodial staff. She was hired as a research assistant in the Department of Genetics, and retired after 30 years as head genetics technician. She is one of my greatest inspirations and I thank her and so many others for breaking down barriers so that I can be where I am today.

CHAPTER ONE: INTRODUCTION

Toward a Better Understanding of Gametocyte Biology and Epidemiology in the Era of Malaria Eradication

1.1 The Malaria Eradication Era and the Shift to Transmission-focused Research

1.1.1 Global Burden of *Plasmodium falciparum* Malaria and the Goal of Eradication

Malaria is a disease with enormous global impact, primarily affecting children in sub-Saharan Africa. The deadliest parasitic disease of humans, there were an estimated 219 million cases of malaria in 2010 resulting in 660,000 deaths [1]. Infection by *Plasmodium falciparum*, the species responsible for the vast majority of these deaths, is characterized by spiking fevers and anemia due to the parasite's replication within and rupture of red blood cells (erythrocytes). Approximately 35% of the world's population lives in areas at risk of transmission with *P. falciparum* [2].*P. falciparum* is transmitted through a mosquito vector and has a very high reproductive index: it is estimated that on average, over 100 secondary cases arise from a single case [3].

Since 2000, due to a worldwide concerted effort to reduce the burden of malaria, deaths from this disease have fallen by 33% in the WHO African Region [1]. This reduction was achieved through a combination of strategies, including the widespread distribution of interventions to block transmission from the mosquito vector (insecticide-treated bed nets and

indoor spraying) and to diagnose and treat more cases (rapid diagnostic tests and first line antimalarial combination therapies) [1]. In 2007, the Bill and Melinda Gates Foundation announced that a major goal of their organization would be to eradicate malaria. The following year, the Malaria Eradication Research Agenda (malERA) initiative was designed to define a specific set of strategies required to achieve successful eradication of this parasitic disease. This is not, however, the first time that malaria eradication has been on the global health agenda. The World Health Organization set out to achieve this goal in the 1950's and initiated the Global Malaria Eradication Program (GMEP). While the GMEP succeeded in eliminating malaria from many regions of the world, in the areas where it failed to eliminate the disease, malaria resurged and eliminated gains that had been made through control efforts. Therefore, in this new era of malaria eradiation, the malERA report concluded that eradication is only possible through the generation of new tools and knowledge to surmount the vast number of challenges on the road to eradication [4].

1.1.2 Malaria's Complex Life Cycle: Challenges and Opportunities for Intervention

Plasmodium falciparum, the deadliest of the human malaria parasites, has a complex life cycle involving growth and replication within two hosts: human and mosquito (See Figure 1.1). *P. falciparum* is transmitted to a human through the infectious bite from a female *Anopheles* mosquito, resulting in the injection of a malaria life cycle stage called the sporozoite. The motile sporozoite homes to the liver, invades a hepatocyte, replicates and produces thousands of progeny, termed merozoites, that enter the host's bloodstream. In the bloodstream, asexual replication begins with merozoite invading an erythrocyte and growing into a young parasite, termed "ring" that develops into a larger trophozoite stage and finally a segmented multinucleated schizont, which ruptures to release 16-32 progeny. Asexual replication causes

damage to the host's erythrocytes and vasculature, resulting in fevers, anemia and coma. Nevertheless, these stages lack transmission potential, and are therefore a dead end in the parasite's life cycle. During the rounds of asexual replication within erythrocytes, a parasite may diverge down a path toward sexual replication. It has been estimated that on average, one sexual stage (termed gametocyte) is produced per 156 asexual stages, thus gametocyte proportion is very low in proportion to the total parasitemia [5]. Gametocytes cause no known damage to the human host, but are importantly the only stages of the parasite's life cycle in the human host that can be transmitted to the mosquito vector. After ingestion by a mosquito during a blood meal, gametocytes burst out of the human erythrocyte, and the female gametocyte further develops into a macrogamete, while the male gametocyte undergoes a process called exflagellation, resulting in the productiong of 8 microgametes. The microgamete penetrates the macrogamete, forming a diploid motile zygote (ookinete), which invades the midgut epithelium. The ookinete grows and divides as an oocyst, subsequently releasing thousands of sporozoites into the midgut. The sporozoites migrate to the mosquito's salivary glands, ready to then be injected into the next human host during a subsequent blood meal.

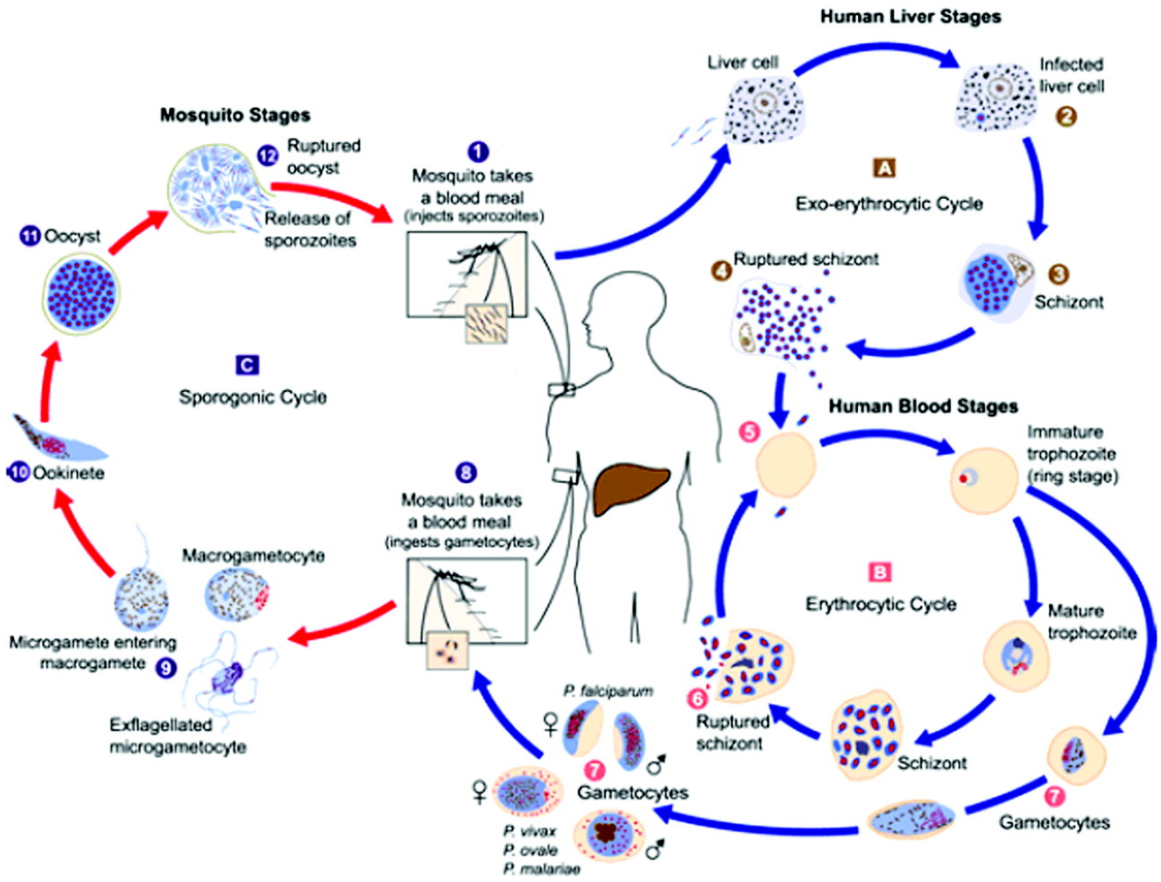


Figure 1.1. The Life Cycle of *Plasmodium falciparum* in the human and *Anopheles* mosquito. Reprinted with permission from [6].

This complex life cycle provides both challenges and opportunities in the pursuit of eradication. Such complexity demands that intervention strategies be varied in their approaches, targeting both the human and the mosquito, both the illness and its transmission. Yet, this complexity also offers opportunities for identifying and exploiting weak points in this complex life cycle. Just 10-100 sporozoites need to be injected into the human host in order to result in a clinically symptomatic malaria infection. Likewise, just one male and one female gametocyte need to be picked up in a blood meal to fulfill the transmission potential of the infection. These gateway stages represent interesting opportunities for interventions aimed at interrupting this complex life cycle.

1.1.3 Shifting Focus: From Morbidity and Mortality to Transmission

The sexual blood stage of malaria has historically been an understudied area of malaria research due to its lack of clinical relevance. Malaria control programs have traditionally focused on reducing morbidity and mortality, and therefore, a main focus of malaria research has been on the asexual stage of the disease. While the asexual stage is associated with the majority of the morbidity and mortality of the disease, this stage has no transmission ability.

The malERA agenda highlights a shift toward blocking transmission and reducing the incidence of new infections. This paradigm shift has placed the gametocyte research as a priority area for knowledge and tool development. In particular, the malERA agenda calls for the development of new tools targeted at the infectious reservoir of gametocyte carriers, and underlines current needs such as a safe and effective drug that kills gametocytes, and a robust field diagnostic for infectious gametocyte carriers [7].

1.2 Gametocyte Development Within the Human Host: The Path to Successful Transmission

1.2.1 Gametocytogenesis

Though the pathologic features of malaria - anemia, fevers, and coma - are attributed to the asexual-replicating stages, transmission of this disease is caused exclusively by the sexual stage, known as the gametocyte. Gametocytes first appear in the peripheral blood approximately one week after parasites are first observed in the blood. Sexual development occurs over a time frame of 8 to 10 days, four to five times longer than the 48 hour development of asexual stages; the lag time before observing gametocytes corresponding to the lag time before the gametocytes reach maturity. As gametocytes develop, they elongate within the host erythrocyte and progress through five morphologically identifiable stages. The end result is the formation of a mature male or female gametocyte. Importantly, these mature gametocyte stages are the only stages that can be successfully transmitted back to the mosquito vector and continue the parasite life-cycle. Prior to their maturity, sexual stages are not present in the peripheral blood [8], suggesting that they could sequester in deep tissue during the week-long development.

1.2.2 Sequestration and Cytoadherence

A hallmark of *Plasmodium falciparum* is the adherence of parasitized erythrocytes to endothelial cells in the microvasculature, a process referred to as cytoadherence. Sequestration of large numbers of infected erythrocytes can result in vessel obstruction, leading to various pathologies, most notably coma and death in the case of cerebral malaria. First described in the late nineteenth century, cytoadherence offered an explanation for the presence of only young asexual “ring” stage parasites and the apparent lack of late asexual trophozoite and schizont

parasites in the circulating bloodstream [9]. In the asexual cycle, the process of cytoadherence is initiated midway through the cycle, such that the young ring stage is found in the blood circulation while the older trophozoite/schizont stages are found in the vasculature of deep tissues. Cytoadherence is a parasite adaptation to evade immune clearance: it hinders the body's ability to eliminate infected erythrocytes as it prevents them from being cleared from the circulation by the spleen. In the spleen, erythrocytes infected with older parasites have an increased rigidity compared with uninfected erythroytes and are therefore retained and destroyed [10].

A number of autopsy studies have been conducted on cases of fatal malaria to explore the distribution of sequestered parasites in the body. A microscopy-based study from postmortem samples determined that the highest degree of sequestration (as defined by percent of erythrocytes parasitized) was the brain, followed by, the heart, liver, lung, kidney and skin, in descending order. All organs surveyed had higher percentages of parasitized erythrocytes –as a percent of total erythrocytes—than the peripheral blood. This finding was later confirmed by studies in which an abundant parasite protein (*Plasmodium* lactate dehydrogenase, pLDH) was assayed across tissues by enzyme-linked immunosorbent assay (ELISA) on tissue homogenates [11], and by immunohistochemistry (IHC) on tissue sections [12]. In the ELISA study, the brain, intestine and skin were observed to have the highest levels of pLDH in the tissue homogenates, with kidney, liver, intestine and lung also positive for pLDH [11]. And in the IHC study, the heart and brain ware observed to have the highest levels of parasitized erythrocytes in tissue sections, with detection also observed in the spleen, liver and kidney [12]. In these studies, as well as ultrastructural analyses of the brain [13,14] and kidney [15], large numbers of parasites

have been observed in contact with endothelial walls, filling the inside of capillaries and venules [13].

Cytoadherence in asexual parasites is achieved through a complex process of host remodeling in which the parasite successfully targets parasite-derived proteins to the surface of the host erythrocyte. PfEMP1 (*P. falciparum* Erythrocyte Membrane Protein 1) is the major parasite ligand involved in cytoadherence and various human endothelial surface proteins, such as CD36 and intercellular adhesion molecule 1 (ICAM-1), have been identified as receptors of PfEMP1 [16]. PfEMP1 is encoded by a family of about 60 different genes called *vars*, that are variantly expressed, and different PfEMP1 proteins have been shown to exhibit different binding properties [17]. The adhesive property of each variant comes from the variable number and combination of duffy-binding like (DBL) domains and cysteine-rich interdomain regions (CIDR) in the large ectodomain of each of the 60 variants [18]. PfEMP1 expression therefore enables sequestration to occur in a tissue-specific manner, due to the differential expression of host receptors across tissues that could interact with PfEMP1 variants. A well characterized example is that of placental malaria: the PfEMP1 protein encoded by the gene *var2csa* has a strong affinity for chondroitin sulfate A (CSA), a protein expressed on the endothelium of the placenta [19].

A number of other proteins are involved in the remodeling of an infected erythrocyte and cause a visible transformation of the cell, as seen by electron microscopy, and result from the translocation of numerous parasite-derived proteins to the surface of the host cell. Within the cytoplasm of the infected erythrocyte, membrane compartments called Maurer's Clefts aid in trafficking of these proteins to the host cell surface. Electron-dense protrusions, termed knobs, emerge on the surface of erythrocytes as the parasite approaches the latter half of its asexual

development [20]. Knobs are hypothesized to aid in adherence due to their positive charge; which could facilitate the interaction of PfEMP1 and its receptor on this host cell since both host cell membranes are negatively charged [21]. In the absence of knobs, adherence is still possible via PfEMP1 in static (non-physiological flow) conditions, however this adherence does not occur under conditions of physiological flow. The gene encoding KAHRP (knob-associated histidine rich protein) has been shown to be required for knob formation, as a genetic knockout resulted in loss of knobs [22].

1.2.3 Sequestration of the Gametocyte

Like the asexual trophozoite and schizont stages, immature gametocytes (stages I, II, III, and IV) are also absent from the blood circulation and therefore hypothesized to sequester through similar mechanisms as the well-characterized process of cytoadherence. However, neither the anatomical localization nor the mechanism of their sequestration is well understood. Important distinctions between asexual and sexual development would suggest that sequestration may involve different mechanisms in the two pathways. In the case of asexual development: the late asexual trophozoite/schizont stage cytoadheres to a vessel for approximately 24 hours during the second half of its development while it replicates its DNA and begins segmenting into daughter cells. These daughter cells then rupture the host erythrocyte to be released as free-living merozoites that subsequently infect other erythrocytes. On the contrary, the sexual stage is sequestered for the majority of its development and enters the bloodstream while still inside the host erythrocyte. Therefore, an important distinction between these processes is that asexual sequestration is an endpoint (ruptured erythrocyte "ghosts" can still be found bound to the endothelium, after the progeny merozoites are long gone), while sexual sequestration is reversible, resulting in the release of the mature gametocyte at the end of sequestration.

Based on early autopsy and biopsy-based studies performed to identify gametocyte sequestration sites, another distinction is that gametocytes appear to be enriched in organs in which asexual sequestration does not predominate. These early studies found that immature gametocytes were enriched in the bone marrow and spleen, as compared with other organs [9,23,24]. This work was later revisited in a study in which bone marrow biopsy blood was compared with peripheral blood. This study revealed an enrichment of immature stages in the bone marrow, despite asexual parasitemia and mature gametocytemia being constant at both sites [25]. Since then, a number of case reports from bone marrow biopsy material have also confirmed the presence of immature gametocytes in this organ [26,27], and the absence of gametocytes in tissues with large asexual parasite burden, i.e. the placenta [28].

Gametocytes have been shown to lack a number of critical components of the well-described asexual cytoadherence machinery. Until recently, controversy existed as to whether the early gametocyte exhibited different cytoadherence-related properties than the intermediate gametocyte. Early reports suggested that all gametocyte stages lacked knobs [29], while another report argued that early gametocytes have knobs but lose them during development [30]. Similarly, there was no conclusion about the binding ability of gametocytes. A number of *in vitro* studies have concluded that young gametocytes bind to cells expressing asexual parasite receptors (i.e. CD36) at far lower levels than asexual stage parasites [31,32,33], while one study concluded that early gametocytes bind at similar levels to asexual parasites [30]. The most likely reason for these discrepancies was the lack of molecular tools and approaches for identifying and purifying gametocytes, and thus the lack of confidence of truly identifying young gametocytes (which are morphologically similar in appearance to asexual stages). Recently, a novel protocol for purifying early gametocytes has allowed investigators to revisit this issue [34]. It has now

been shown that all gametocytes lack knobs as well as expression of the required knob-associated histidine rich protein (KAHRP). Additionally, very low expression levels of *var* genes and PfEMP1 protein were found in gametocytes [33,35].

However, it is still unknown whether gametocytes have the ability to truly cytoadhere. A number of studies have shown little to no binding of gametocytes to CD36 and ICAM-1 [30,31,33]. Two studies were performed using bone marrow endothelial cells, under the hypothesis that gametocytes sequester in the bone marrow, and observed higher levels of binding. Through testing a panel of antibodies to block the binding, the authors identified ICAM-1, CD49c, CD166 and CD164 as candidate receptors of gametocyte sequestration in the bone marrow [36]. However, these receptors have yet to be confirmed in any subsequent studies. A subsequent study that compared binding of stage III-IV gametocytes to three endothelial cell lines showed that the highest rate of gametocyte binding was observed in the human bone marrow endothelial cell (HMBEC) line, although the rate was still lower than that of asexual cells [31]. Both authors note that while gametocytes appear to bind at some low level to bone marrow endothelial cells, the binding efficiency observed in these assays is not sufficient to explain the depletion of all immature sexual stages from the bloodstream.

As it appears unlikely that gametocytes utilize the classical method of cytoadherence employed by asexual stages, new hypotheses have arisen suggesting that gametocytes could employ a mechanical-based method to remain sequestered during development. As they progress through stages I through V, gametocytes undergo a dramatic shift in deformability, becoming quite rigid during development (peak stiffness at stage IV), and subsequently very deformable upon maturity[37,38,39]. It has been hypothesized that these properties could allow the gametocyte to remain stuck and mechanically sequestered in the spleen, and then deform

enough to get back into the circulation upon maturity, when the gametocyte needs to be picked up by a mosquito to continue the cycle. Mathematical modeling has demonstrated that based on their dimensions and deformability characteristics, this hypothesis is theoretically possible [38], however, it remains to be observed *in vivo*.

Many questions remain in the pursuit of understanding gametocyte sequestration: i) where exactly do immature gametocytes reside in the human host while they develop?, ii) why do they sequester in distinct anatomical niches compared to asexual parasites?, iii) through what mechanism are gametocytes staying sequestered from the bloodstream during their development, finally, iv) through what mechanism do they return to the blood stream after maturation?

1.2.4 Significance of Understanding Gametocyte Development in the Era of Malaria Eradication

Elucidating uncharacterized processes in gametocyte development may unveil new targets for therapeutic intervention. Drugs that effectively kill gametocytes are predicted to have the best impact in reducing new cases of malaria in low transmission settings, where infections occur more rarely population-wide and thus prevention of each new infection is critical [40]. Thus, as the number of malaria cases continues to fall, a drug that effectively kills gametocytes will be an important tool for malaria eradication.

Drugs that target the gametocyte are lacking as this has historically been a neglected area for malaria drug development efforts as most drugs are designed to cure disease symptoms, not necessarily to prevent transmission. Artemesinin (the primary component of most first line antimalarials employed globally for *P. falciparum*) and its derivatives have been shown to be ineffective in killing mature gametocytes *in vitro* [41,42] and 5-39% of artemesinin-treated

individuals have been found to carry gametocytes one month after completing their treatment course [43,44]. The only currently licensed antimalarial drug with clear efficacy against gametocytes is primaquine [45], however, this drug has been linked to a dangerous side effect in individuals with glucose-6-phosphate dehydrogenase (G6PD) deficiency, a blood condition most common in parts of the world with a high malaria burden [46].

1.3 Gametocyte Detection: Understanding the Infectious Reservoir

1.3.1 Gametocyte Detection

In order for malaria to be transmitted to the mosquito vector, a mosquito must ingest at least one male and one female gametocyte from a blood meal. Therefore, the presence of mature gametocytes in the peripheral blood has historically been used as an indication of an individual's potential infectiousness. The standard method for detecting gametocytes is the identification of crescent-shaped mature gametocytes in a thick smear examination; however, this approach is limited by its low sensitivity. As low as 1-10 gametocytes per uL of blood can lead to infection, and thick smear sensitivity is only in the range of 10-100 gametocytes per uL thus missing individuals with low level gametocyte burden [47]. To increase sensitivity of gametocyte detection, transcriptional approaches such as reverse transcriptase PCR (RT-PCR), quantitative nucleic acid sequence based amplification (QT-NASBA) and reverse transcription loop-mediated isothermal amplification (RT-LAMP) have been developed for gametocyte detection using established molecular markers of gametocyte development [48,49,50].

Molecular methods for detecting gametocytes have been shown to be 100-1000 times more sensitive than thick smear [47,51], and thus have painted a very different picture of

gametocyte carriage than what was previously established based on microscopy alone. Population surveys using the mature gametocyte marker *Pfs25* in QT-NASBA have shown that between 15-70% of cross-sectional endemic area study cohorts [48,52,53] and 39-89% of uncomplicated malaria cohorts are *Pfs25*-positive [51,54,55,56,57,58], revealing a much larger reservoir of submicroscopic gametocyte carriers than previously hypothesized.

1.3.2 The Infectious Reservoir

While children under five years of age in endemic settings account for the majority of malaria infections, and suffer the greatest morbidity and mortality of the disease [1], the adult population remains an important factor in the context of the epidemiology of gametocyte carriage. A collection of cross-sectional surveys of gametocyte carriage in Burkina Faso revealed that an individual's gametocyte proportion increases with the age of the host. By microscopy, it was estimated that 1.8% of parasites in the bloodstream of young children are gametocytes versus 18.2% in adults, a striking difference also confirmed by QT-NASBA [59]. This study confirmed earlier data from endemic settings in Tanzania and the Gambia suggesting that adults contribute substantially to the infectious reservoir [60]. These studies measured the rate of infectious gametocyte carriers in each age group, and adjusting by the population age structure, concluded that between 28-38% of all transmission in these endemic settings comes from adults. Further, there is a temporal dynamic to gametocyte carriage over the course of one's infection. Certain physiological factors are believed to induce a higher fraction of parasites to diverge and proceed down a path of sexual development and therefore increase infectiousness [61]. Anemia and long illness duration were found to be associated with higher gametocyte carriage in a cohort of patients from Thailand [62].

1.3.3 Significance of Improving Gametocyte Detection in the Era of Malaria Eradication

Molecular methods have made it feasible to detect gametocytes within the entire range that is believed to be biologically relevant for infectivity. Interestingly, these methods have revealed that a much larger fraction of infected individuals harbor gametocytes than previously thought. If in fact all infections harbor gametocytes to some degree, then there is a need for more sophisticated assays to characterize these gametocytes. Although mature gametocytes are required for infectivity, mosquito feeding studies have shown that there is a weak correlation between gametocyte density and infectivity, and an elevated level of gametocytes does not necessarily always translate to higher levels of infectivity [63]. Therefore, studies are required to determine what factors separate infectious gametocytes from non-infectious gametocytes and whether there are better markers for infectivity. New molecular tools could be used to identify individuals who are infectious (which could depend on age of gametocyte, sex ratio of gametocytes, or other marker of health of gametocytes) or who carry the young gametocyte ring stage, which is hypothesized to circulate in the peripheral blood for a short time (less than 24 hours) prior to sequestration [25]. This young gametocyte could potentially serve as an early biomarker for conversion to sexual development. Such a tool would be of great aid in understanding triggers of gametocytogenesis *in vivo*.

The previous malaria eradication program lacked molecular methods for careful monitoring of gametocyte carriage, however real-time monitoring of gametocyte carriage and infectivity is important in pre-elimination settings. Malaria is a dynamic ecological system and therefore when its population size is reduced, there may be selection for highly infectious strains. There is some evidence to suggest resurgences are in part due to selection for high gametocyte-producing strains [64]. In a resurgence that occurred in 1960 in Indonesia where control

programs had reduced but not eliminated malaria, an increased number of gametocyte carriers was noted [65]. And in 1980, over the course of the Garki Project (a malaria control study carried out in Northern Nigeria) gametocyte carriage among 9-28 year olds showed a significant increase compared to that of the same age group in a control village with no intervention in the same time frame [66].

1.4 Summary of Aims

This dissertation aims to contribute to the field's basic knowledge of gametocyte biology as well as to provide tools for detecting gametocytes. Thus, the first aim is to characterize the process of gametocyte sequestration as there has yet to be a comprehensive study on this topic in the human host using molecular approaches. Using samples from autopsy cases of fatal pediatic malaria in Malawi, we employ a combination of immunohistochemistry (IHC) and qRT-PCR to identify enrichment sites of immature gametocytes in the human host. Further, we characterize tissue-specific patterns of gametocyte sequestration using IHC to co-localize gametocytes with host proteins. Finally, we use electron microscopy to examine the ultrastructure of sequestering gametocytes.

In the second aim, we develop a framework and assay for identifying young, intermediate and mature gametocytes from within a mixed population of parasites. At any point during infection with *P. falciparum*, a peripheral blood sample is likely to contain a mixture of asexual and sexual stages. We re-analyze existing microarray data to identify stage-specific signatures within a mixed stage population. We then use these signatures to define a small set of sentinel genes that could be used to infer the stage composition of a sample. Finally, we apply the system

to identify carriers of young and mature gametocytes in cohorts of severe and uncomplicated malaria patients.

CHAPTER TWO

Evidence for Malaria Sexual Stage Development in the Hematopoietic System of the Human Bone Marrow

Regina Joice[1], Jacqui Montgomery[2*], Sandra K. Nilsson[1*], Danny A. Milner[1,3], Selasi Dankwa[1], Elizabeth Egan[1], Belinda Morahan[4], Karl B. Seydel[5,6], Kim C. Williamson[4], Manoj T. Duraisingh[1], Terrie E. Taylor[5,6], Matthias Marti[1]

[1]Harvard School of Public Health, Department of Immunology and Infectious Diseases, 665 Huntington Ave, Boston MA 02115

[2]Malawi-Liverpool-Wellcome Clinical Research Programme, PO Box 30096, Chichiri, Blantyre 3, Malawi

[3]Brigham and Women's Hospital, Department of Pathology, 75 Francis Street, Boston MA 02115

[4]Loyola University, Department of Biology, 1032 W. Sheridan Road, Chicago IL 60660

[5]Blantyre Malaria Project, University of Malawi College of Medicine, Blantyre, Malawi

[6]Michigan State University, College of Osteopathic Medicine, East Lansing MI 48824

Author contributions: **R.J**., **S.K.N**., **D.A.M.** and **M.M.** designed the research; **R.J.**, **S.K.N., S.D., E.E.** and **B.M.** performed the research (**R.J.**: *in vitro* time course experiments, fluorescence microscopy, IHC antibody optimization, IHC organ screen, qPCR primer optimization, qPCR organ screen, ImageJ cell size analysis, IHC double-labeling, electon microscopy, **S.K.N., S.D., E.E**: erythrocyte precursor assays, **B.M**.: IHC antibody optimization); **J.M., D.A.M., K.S.** and **T.E.T.** organized and performed sample collection; **K.C.W.** and **M.T.D.** contributed helpful discussion; and **R.J., S.K.N.** and **M.M**. wrote the manuscript.

Abstract

In the current study, we systematically quantified gametocyte sequestration in autopsy cases from an ongoing study of fatal pediatric malaria in Blantyre, Malawi. We performed an organ survey using immunohistochemistry (IHC) on tissue sections from nine body sites (brain, lung, heart, intestine, liver, kidney, subcutaneous fat, spleen, and bone marrow), which suggested an enrichment of gametocytes in the bone marrow. Quantitative RT-PCR supported this finding and confirmed the presence of young developing gametocytes in this organ. Expanding our survey to thirty patients, we performed double-labeling of parasite and host markers to sub-localize gametocytes within this organ. We observed gametocytes enriched in the extravascular space of the bone marrow; frequently co-localized with erythroblastic islands. We show that gametocytes are infrequently found within bone marrow macrophages, and that by electron microscopy, we observed that a subset of parasitized cells in the bone marrow appear knobless, both in agreement with recent studies suggesting that gametocytes do not alter the surface of the host erythrocyte as is observed with asexual stages. Finally, we show both *in vivo* and *in vitro*, using a model system of erythropoietic development, that gametocyte development can occur within erythrocyte precursor cells (orthochromatic erythroblasts and reticulocytes). Understanding the process through which *P. falciparum* gametocytes sequester outside of the bloodstream during development is of critical importance, considering the field´s renewed focus on understanding the dynamics of malaria transmission and the development of new strategies to interrupt it.

Introduction

Sequestration of asexual stage infected red blood cells is a major means of severe pathogenesis in falciparum malaria [67], but it is not known to what degree gametocytes share mechanisms of sequestration. The mechanisms of tissue-specific sequestration of asexual stages have been studied in great detail and revealed the molecular processes of host cell manipulation and subversion by the parasite [68,69,70]. Despite the importance of sequestration for survival during the long period of sexual development, little is known about the underlying molecular mechanisms and sites of tissue sequestration for these stages. In autopsy case studies performed in the 1930's [23], gametocytes were observed in the bone marrow and the spleen. This early work was revisited through the study of bone marrow biopsy samples, which also revealed the presence of immature gametocytes in that organ [24,25,26]. Many attempts have been made to demonstrate *in vitro* binding of gametocytes to endothelial cells or purified receptors, and expression of known asexual parasite cytoadherence ligands in developing gametocytes. These studies have shown low levels of both binding [30,31,32,36,71] and parasite adhesion ligand expression [33], and thus it has been suggested that different receptors and/or a novel mechanism could be involved in the sequestration of developing gametocytes, distinct from the well-described binding interaction of asexual stages [26,33,72]. In support of this hypothesis it has recently been shown in *P. falciparum* that gametocyte sequestration dynamics coincide with alterations in gametocyte rigidity, suggesting that some form of mechanical sequestration may be possible [38,39]. Here we present the results of a systematic organ survey, using the largest existing collection of autopsy cases in fatal malaria [73], to identify and characterize gametocyte sequestration sites.

Materials and methods

Autopsy cohort

The institutional review boards (IRB) at the University of Malawi College of Medicine, Michigan State University, and the Harvard School of Public Health approved all aspects of autopsy cohort study and the severe malaria peripheral blood cohort study. Informed written consent was obtained from all parents/guardians of the enrolled patients. IRB at the Brigham and Women's Hospital approved the use of discarded surgical tissue for laboratory experiments.

Autopsy cohort

The autopsies analyzed in this study are part of a pediatric severe malaria and postmortem autopsy study that was performed between 1996 and 2011 [73]. Children meeting the clinical case definition of cerebral malaria (CM) were admitted to the Malaria Research Ward, located in the Queen Elizabeth Central Hospital in Blantyre, Malawi and enrolled in the severe malaria study upon the consent of the parent or guardian. Criteria used for diagnosis and clinical management have been previously described [73]. Over 3000 individuals were enrolled in the study over this time period. The mortality rate for the study participants is between 15-20%. Consent was given to conduct an autopsy for 103 cases, and these autopsies were performed between 2 and 14 hours after death.

Tissue specimens were collected in various formats at the time of autopsy. Specimens were stored in 10% buffered formalin to be later used for histological examination and immunohistochemistry analysis. Storage in glutaraldehyde buffer enabled specimens to be processed for electron microscopy. RNAlater snap frozen samples were later used for qRT-PCR analysis.

Severe malaria peripheral blood cohort

The peripheral blood samples analyzed in this study are part of a severe malaria study, using an entire seasonal cohort from January to June of 2010 [73]. Seventy-six children were enrolled that season, with the consent of the parent or guardian. Three milliliters (mL) of venous blood was collected at time of admission in Tri-Reagent BD, shaken for 1 minute, and then frozen and stored at −80°C. A portion of this 3mL sample was used for qRT-PCR analysis.

***Plasmodium falciparum in vitro* culture**

Parasites of the *P. falciparum* reference strain 3D7 and Pf2004 were used for this study. Parasites were maintained in fresh type O+ human erythrocytes (Research Blood Components, LCC) at 2% hematocrit in HEPES-buffered RPMI 1640 containing 1% (w/v) AlbuMAX II (Invitrogen), 0.5 ml Gentamycin, 2.01 g sodium bicarbonate and 0.05 g Hypoxanthine at pH 6.74. Cultures were kept in a controlled environment at 37°C in a gassed chamber at 5% CO_2 and 1% O_2. For gametocyte production for qRT-PCR marker testing, gametocytogenesis was triggered using the 3D7 parasite line. Sorbitol synchronized ring-stage parasites were seeded at 1-2% parasitemia into 75cm^2 vented flasks at 6% hematocrit. Gametocytogenesis was then triggered according to the Fivelman protocol [74]. Briefly, asexual parasites were grown to a high parasitemia in the presence of partially spent ("conditioned") medium, and then sub-cultured at the schizont stage into new dishes containing fresh media and erythrocytes.

Controls for immunohistochemistry

To determine antibody specificity in the tissues of interest, we generated formalin-fixed and paraffin-embedded tissue samples containing erythrocytes infected with asexual and sexual parasites stages. Briefly, 100 µl of packed erythrocytes, derived either from asexual cultures or from individual time points of gametocyte development time courses, were mixed with thrombin and fetal bovine plasma, wrapped in lens paper and fixed with formalin. Cell clots were then placed into a paraffin wax block together with a composite of surgical tissues obtained from the Brigham Hospital from the same 9 organs that are analyzed in the autopsy study, and sliced at 3 µm. Using these samples, pLDH antibodies were titrated for parasite-specific labeling and Pfs16 antibodies both for parasite- and gametocyte-specific labeling by IHC, with tissue samples allowing for antibody titration against tissue-specific background (Appendix Figure A3).

Immunohistochemistry assays

Formalin-fixed paraffin-embedded tissues and control blocks were cut into 3 micron sections and mounted on slides. Sections were dried overnight at 37°C, deparaffinized in xylene, and hydrated through a series of graded alcohols, finishing in water. Antigen retrieval was performed by incubating slides at approximately 95°C in a steamer for 20 minutes. One of two antigen retrieval solutions was used for this process (as determined based on antibody optimization conditions): 1 mM EDTA + 0.05% Tween at pH 8 or Retrieve-All-2 antigen retrieval buffer pH 10 (Covance Research Product, Dedham, MA). EDTA solution was used for the initial pLDH and Pfs16 screen, as well as for the double-labeling of pLDH and Pfs16 with

PECAM-1. For the double-labeling of parasite antibodies with CD163 and CD71/Transferrin Receptor, Retrieve-All-2 buffer was used. Following antigen retrieval, slides were blocked using a universal blocking buffer (Thermo Scientific) for 20 minutes, followed by 10 minutes each of avidin and biotin blocking steps to block endogenous biotin and avidin, respectively. Antibody incubation was then performed for either 2 hours (PECAM-1) or 1 hour (all other antibodies), and each antibody was diluted in blocking buffer. Wash steps, in between each step therafter were performed with TBS + 5% Tween for 3 x 5 minutes.

For the initial screen, secondary anti-mouse biotinylated antibodies were used, followed by streptavidin conjugated to alkaline phosphotase (AP). For the development of signal, naphthol phosphotase-fast red chromagen reagent was applied. Slides were subsequently rinsed in water and counterstained in Mayer's hematoxylin, and mounted in aqueous mounting medium. Slides were blinded to patient ID and independently counted by two microscopists, counting parasites in 100 consecutive high power fields, starting in the upper left corner of each section. The average of the microscopists' counts were used for the analysis.

For double labeling with platelet endothelial cell adhesion molecule (PECAM-1), a sequential double staining protocol was used, as previously described [75], as both antibodies were raised in the same host. Here, primary human host antibody (mouse anti-PECAM-1) was incubated for 1 hour, followed by a peroxidase block (to block endogenous peroxidase activity), secondary anti-mouse biotinylated antibody, streptavidin conjugated to horse radish peroxidase (HRP) and development of signal with 3,3'-Diaminobenzidine (DAB) chromagen reagent. After washing in TBS, blocking steps with avidin and biotin were repeated (to block residual unbound biotin or streptavidin from first reaction). The parasite antibody (mouse anti-pLDH or mouse anti-Pfs16) was incubated for 1 hour, followed by a secondary anti-mouse biotinylated antibody,

streptavidin conjugated to alkaline phosphotase (AP), and development of signal with naphthol phosphotase-fast red chromagen reagent. Slides were subsequently rinsed in water and counterstained in Mayer's hematoxylin, and mounted in aqueous mounting medium.

For the CD163 and CD71 double labeling experiments, we used a double-staining protocol for mouse-rabbit combinations, as described in [75]. Here, a primary antibody for a human host protein (either rabbit anti-CD163 or rabbit anti-CD71) and a parasite protein (rabbit anti-BIP or rabbit anti-Pfs16) were incubated simultaneously. This was followed by a peroxidase blocking step, and the addition of secondary antibodies (biotinylated goat anti-rabbit and HRP-conjugated goat anti-mouse), followed by streptavidin conjugated to alkaline phosphotase (AP). The HRP was developed by a brief incubation with 3,3',5,5'-Tetramethlybenzidine (TMB) substrate and washed with TBS. The AP was then developed by incubation with naphthol phosphotase-fast red chromagen reagent. Slides were subsequently rinsed in water and mounted in aqueous mounting medium.

Parasite and host marker antibodies

Working concentrations for each antibody were determined using the control samples as described above. The initial screen was performed using antibodies targeting pLDH, *Plasmodium* lactate dehydrogenase, which is constitutively expressed in parasites and previously identified as a useful marker of malaria parasites in autopsy tissue [12], and Pfs16, a gametocyte antigen expressed throughout the duration of gametocyte development [76]. Here we used a mouse monoclonal anti-pLDH antibody (kind gift from Dr. Michael Makler, Flow, Portland, OR) at 1:750 and a mouse polyclonal anti-Pfs16 antibody (provided by Dr. Kim Williamson, Loyola University) at 1:5000. The co-localization screen with platelet endothelial cell adhesion

molecule (PECAM-1), a marker of humam endothelial cells, was performed with these same parasite antibodies as well as a mouse anti-PECAM-1 antibody (Clone JC70A, Dako) at 1:1000. The co-localization of parasite proteins with CD163, a marker of macrophages, and CD71, expressed on erythroblasts and reticulocytes, was performed using Pfs16 and BIP, a heat shock protein expressed in all blood stages. In this case, due to the need to co-localize parasites within host cells, we optimized protocols for rabbit parasite antibodies and human host mouse antibodies to minimize cross-reactivity. We generated a rabbit anti-Pfs16 antibody against a peptide of Pfs16 for this purpose. This rabbit Pfs16 antibody was used at 1:4000 and the rabbit anti-BIP antibody (provided by MR4, clone MRA19) was used at 1:750. Host proteins were labeled using mouse anti-CD163 (Leica, clone 10D6) at 1:25 and mouse anti-CD71 (Invitrogen, clone H68.4) at 1:1000.

Autopsy tissues used in each analysis

For the initial screen, we first performed IHC using platelet endothelial cell adhesion molecule (PECAM-1), a marker of host endothelial vessels on bone marrow tissue from the 21 most recent autopsy cases. This was done as a quality control measure for our subsequent tissue analysis. Of those 21, PECAM-1 labeling failed in 3 individuals, which we then excluded from further analysis. Among these 3 excluded individuals were the 2 individuals with the longest post-mortem intervals (over 12 hours in length). We also excluded HIV positive patients from our analysis because the effect of HIV on the natural progression of malaria infection is not well understood [27]. This excluded 8 additional patients. This left us with ten HIV-negative, PECAM-1-positive autopsy cases for which we performed our organ screen. These cases had post-mortem intervals ranging from 4 to 12 hours.

For the subsequent characterization of parasites intra- and extravascularly using double labeling with PECAM-1, we expand our cohort. From the remaining 82 cases in the autopsy study that we did not use for the initial organ screen, we identified a subset of 28 patients based on applying the following filters: (a) HIV-negative status, (b) post-mortem interval under 9 hours, and (c) availability of bone marrow FFPE block. Six of these cases were subsequently excluded from further analysis due to the lack of PECAM-1 staining. For the remaining 22 bone marrows, 20 had pLDH-positive parasites and 16 had Pfs16-positive parasites. These 22 in addition to the 8 cases from the initial screen in which gametocytes were found in the bone marrow gave us a total of 30 patients upon which to examine intra- and extravascular bone marrow parasite localization.

For the co-localization of parasite markers with CD163 and CD71, we used the 22 bone marrows in which at least 5 pLDH-labeled parasites were detected in 50 high power fields in the previous screen. For these 22, we applied double-labeling of both CD163 and CD71. All 22 labeled successfully with CD163 and CD71. For the co-localization of Pfs16 with CD71, data is plotted for only the 8 patients in which at least 5 Pfs16-labeled parasites were detected in 50 high power fields.

Size Measurements with ImageJ

Images of parasites within control *in vitro* blocks and human autopsy tissue were taken using a digital microscopy camera (Zeiss AxioCam IC). These images were processed using ImageJ software. In ImageJ, a threshold was applied to each image to create masks of each positively-labeled cell in a binary black and white format. The Analyze Particles feature was subsequently applied to each mask, and measurements of area, diameter, perimeter and circularity in microns

were recorded for each particle.

Immunofluoresence

For the evaluation of quantitative PCR markers and for localization of gametocytes in erythrocyte precursor cells, immunofluoresence assays were performed with cell monolayers on glass slides, prepared as described previously [77]. For the purpose of evaluating PCR markers, slides were fixed in ice-cold methanol, blocked with 5% nonfat dry milk powder, incubated with polyclonal mouse antibody against Pfs16 (1:2500) [78], washed and incubated with a secondary antibody conjugated to Alexa 488. Parasite nuclei were labeled with DAPI and quantification was done on the proportion of FITC(+) parasites out of the total number of DAPI(+) parasites. For erythrocyte precursor cell localization, slides were fixed with 4% paraformaldehyde and 0.0075% glutaraldehyde. Slides were subsequently washed in PBS and incubated for 10 min with wheat germ agglutinin (WGA) directly conjugated to Alexa 488 (8 μg/mL, Invitrogen). After washing, cells were permeabilized with 0.1% Triton X-100 in PBS for 10 min, blocked with 3% skim milk in PBS for 30 min, and incubated with rabbit anti-Pfs16 (1:40,000) for 1h. Slides were washed and primary antibodies detected by incubation for 1h with Alexa-594-conjugated anti-rabbit IgG (1:500, Invitrogen). All WGA and antibody dilutions were made in PBS supplemented with 3% bovine serum albumin. To avoid photobleaching, cells were mounted with Vectashield (Vector Laboratories) containing 1.5 mg/mL DAPI to stain for DNA. All incubations were performed at room temperature in a humid chamber.

Transmission Electron Microscopy

Transmission Electron Microscopy (EM) was performed on the bone marrow and brain from one

patient for which we had well-preserved tissue for EM, and both parasites and gametocytes were identified in the patient's brain and bone marrow during the initial IHC screen. Tissue taken at autopsy was fixed in 2% glutaraldehyde. Samples were then post-fixed in osmium tetroxide, dehydrated in ethanol, treated with prophylene oxide and emdedded in Epon-Araldite. Thin sections were stained with 1% toluidine blue and examined under light microscopy to identify sections of the tissue that contained parasites. Ultra-thin sections were then made and stained with uranyl acetate and lead citrate just prior to imaging. Bone marrow sections were imaged using a Tecnai G2 Spirit BioTWIN conventional electron microscope. Brain sections were imaged using a JEOL 1200EX- 80kV conventional electron microscope.

RNA Extraction, DNAse Digest and Reverse Transcription.

Tissue sections stored in RNALater were processed through either grinding in liquid nitrogen using a mortar and pestle or homogenizing in TriReagent BD (Molecular Research Center). All TriReagent BD samples were then processed as previously described [79]. Briefly, RNA was extracted by the phenol-chloroform method followed by DNAse digest (Ambion), and a second phenol-chloroform extraction for protein removal and sample concentration. First strand cDNA synthesis was performed using the SuperScript III First Strand Synthesis kit (Invitrogen).

qRT-PCR assay

To develop a qRT-PCR assay capable of differentiating between developmental gametocyte stages, we selected three established markers with peak times during early (*PF14_0748*), mid (*Pf4845*) and late (*Pfs25*) gametocyte development, respectively. To establish stage-specificity a series of *in vitro* gametocyte time courses were performed using the protocol according to

Fivelman et al [74], and samples were collected daily during asexual growth and subsequent gametocyte development. In addition, we determined the sensitivity of each primer pair in dilution curves using plasmids containing the cloned target fragments as template. These experiments demonstrated a detection limit ranging between $10^1 – 10^2$ copies for each primer pair. All primer pairs were shown to be not cross-reactive with human cells using cDNA from human C32 cells.

qRT-PCR conditions

Quantitative PCR was performed on cDNA in 20 uL reaction volumes in either an ABI 7300 or Viaa7 machine (both, Applied Biosystems). Amplification was performed using iQ SYBR Green Master Mix (BioRad), and primers for each of the genes of interest at 250nM concentration. Reaction conditions were as follows: 1 cycle x 10 min at 95°C and 40 cycles x 30 sec at 95°C and 1 min at 58°C. Each cDNA sample was run in triplicate, and each plate contained a dilution series of cDNA amplicons cloned into pGEM plasmids. As a positive control to ensure parasite RNA was present in our samples, we used primer pairs for *P. falciparum* ubiquitin conjugating enzyme transcript as a constitutively expressed parasite marker as described previously [80].

Erythrocyte precursor cell culture

CD34+ hematopoietic stem cells isolated from a single donor were purchased commercially. Cells were co-incubated on a stromal cell monolayer. On day 19 of culture, cultured red blood cells (cRBCs) were collected for parasite growth assays (41% of the cRBC were nucleated and 59% were enucleated, and of the latter 95% were reticulocytes and 5% were normocytes as

determined by New Methylene Blue staining). Parasites of the strain Pf2004 were sorbitol-synchronized at the ring stage. They were then induced to form gametocytes as described previously [74] until reaching the late trophozoite/schizont-stage. At this point, infected erythrocytes were enriched by magnetic cell sorting (MACS), which yielded a parasitemia of 92%. Parasites were plated with cRBCs in triplicates at a final parasitemia of 1.1% and 0.7% hematocrit in 100 µL RPMI supplemented with 10% human AB+ serum. Following reinvasion, cells were collected by cytospin and stained with Giemsa to monitor parasite growth. Media was changed daily until day 4. On day 1 and 4, cells were collected by cytospin onto poly-L-lysine coated slides for immunofluorescence microscopy assays (IFA), and slides were air-dried and stored desiccated at -20°C.

Results

Organ Survey to Determine Enrichment Sites of Immature Gametocytes.

Immunohistochemistry-based organ survey. We applied an immunohistochemistry-based labeling approach to autopsy tissue sections from cases of clinically diagnosed cerebral malaria. *P. falciparum* lactate dehydrogenase (pLDH) was used to label all parasites [12] and the gametocyte antigen Pfs16 [78] to label all gametocytes. In the initial screen, we surveyed tissue from eight organs plus subcutaneous fat from a total of six autopsy cases. This screen revealed the highest levels of gametocytes in the spleen, brain and bone marrow (Figure 2.1A/B and Appendix Figure A1). To further investigate gametocyte loads in the three candidate organs we included samples from 4 additional patients. In this 10 patient cohort, 8 had detectable gametocytes in at least one tissue. The spleen, brain, heart and gut revealed the highest total parasite densities (Figure 2.1B), which is consistent with previous studies showing high levels of asexual parasite sequestration in these organs [11,13,81]. To normalize differences in total parasitemia due to the heterogeneity in vascularity between tissues, we assessed gametocyte enrichment as the number of Pfs16-positive cells divided by the number of pLDH-positive cells in 100 high power fields (HPF) for each sample. The only organ with a mean gametocyte proportion greater than 50% was the bone marrow (Figure 2.1C). All other organs had a mean gametocyte fraction under 25%, and all but spleen were under 13%, which is consistent with previously reported low percentages of circulating gametocytes in human malaria infection [63].

Quantitative PCR-based assessment of gametocyte age across tissues. To determine whether the bone marrow is enriched for the early and intermediate stages of gametocyte development that are known to occur outside of blood circulation, we measured the tissue-specific expression

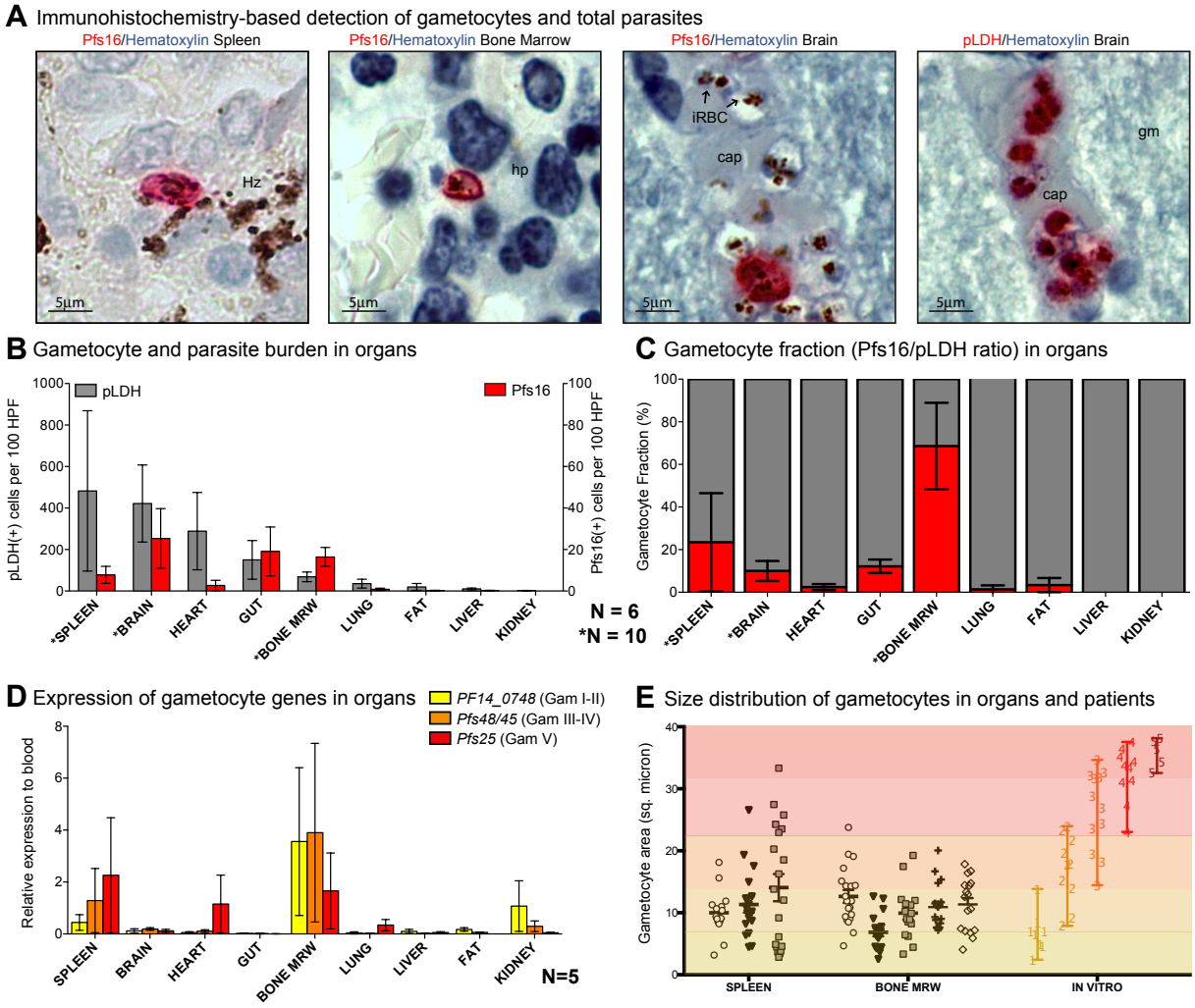
of three genes that have peak expression at different points throughout gametocyte development. Based on microarray evidence and previous validation in the literature [82,83], as well as our own *in vitro* qRT-PCR time course experiments (Appendix Figure A2), we selected the *PHISTa* gene *PF14_0748* as a marker of young gametocytes, transmission-blocking antigen precursor *Pfs4845* for intermediate gametocytes, and ookinete surface antigen precursor *Pfs25* for late developing and mature gametocytes. As we were interested in whether there was an enrichment of young gametocytes in any organ, relative to what is found in the peripheral blood, we aimed to look at gene expression relative to that found in peripheral blood. We were unable to collect peripheral blood from these autopsies, so we instead analyzed peripheral blood from an entire seasonal cohort of severe malaria patients, who were enrolled in the severe malaria same study as the autopsy cases (N= 78). Taking the mean expression level of these three markers across the peripheral blood samples, we determined the fold change enrichment of each gametocyte marker in each organ. This analysis revealed that the expression of the young and intermediate gametocyte markers *PF14_0748* and *Pfs4845* peaked in the bone marrow, while expression of the mature gametocyte marker *Pfs25* was distributed across several organs, including the spleen, heart and bone marrow with low level expression in the brain (Figure 2.1D).

Size-based assessment of gametocyte age across tissues. To further assess age of gametocytes across organs, we measured the size of Pfs16-labeled gametocytes from the 2 tissues with the highest gametocyte fraction (spleen and bone marrow). For this assessment, we used tissues where we could detect at least 15 gametocytes, resulting in 3 spleen and 5 bone marrow sections to use for this analysis, from a total of 5 patients (Figure 2.1E). We performed size measurements using ImageJ on *in vitro* gametocytes of known stage stored and prepared in the same way as the autopsy tissue sections. In each of the five bone marrows examined,

gametocytes were very rarely of a size corresponding to a stage IV-V gametocyte, while in the spleen, large gametocytes were far more common (Figure 2.1E). Generally bone marrow-residing gametocytes across patients are observed to fall within a tight size range, corresponding to stage I, II and III gametocytes from our control *in vitro* blocks. This further illustrates a specific enrichment of young gametocytes in the bone marrow.

Figure 2.1. Enrichment of developing gametocytes in the human bone marrow. (A). Representative images of Pfs16-labeled gametocytes and pLDH-labeled parasites in bone marrow, spleen and brain. Parasite antigens are visualized with alkaline phosphatase (AP) FastRed chromogen and nuclei are labeled with hematoxylin blue stain. The panel shows large gametocytes in the spleen and brain, as well as smaller gametocyte in the bone marrow. Shown are also multiple large pLDH-labeled trophozoite or schizont stage parasites sequestered in a brain capillary. A complete set of representative images of labeled parasites from all tissues is shown in Appendix Figure A1. In the spleen, free hemazoin can be observed. In the bone marrow, parasite appears in context with nucleated cells of the hematopoietic system. In the brain, Pfs16 labels a subset of parasites within capillaries while pLDH labels all parasites. Hz: free hemozoin. hp: hematopoietic cells of the bone marrow. iRBC: unlabeled infected RBC. cap: capillary in the brain. gm: gray matter in the brain. **(B). Quantification of parasite and gametocyte load across tissues.** Numbers of pLDH-positive (left axis) and Pfs16-positive cells (right axis) were quantified per 100 high power fields (HPF) for each tissue. For bone marrow, spleen and brain samples from 10 patients were analyzed, and for the remaining tissues samples from 6 patients were analyzed. The observed distribution confirms previous studies with highest parasite loads in spleen, brain, heart and gut. **(C). Comparison of gametocyte fractions across tissues.** Fractions are calculated as ratio of Pfs16-positive cells to pLDH-positive cells in 100 high power fields from each sample, for each tissue in which at least 10 pLDH parasites were observed. This approach normalizes out the variability in total parasite load across patients and demonstrates highest enrichment of gametocytes in the bone marrow, with a mean fraction of > 60% across 10 patients.

Figure 2.1 (Continued). (D). qRT-PCR analysis to quantify gametocyte stage distribution across tissues. Markers for early developing (Stage I/II, *PF14_0748*), mid developing (Stage III/IV, *Pfs4845*), and late developing and mature gametocytes (Stage IV/V, *Pfs25*) were used for gametocyte quantification. RNA from each of the nine tissues from 5 patients were analyzed, as well as RNA from the peripheral blood of 78 patients enrolled during one season of the severe malaria study were analyzed. The data are shown as relative gene expression (fold increase) of each of the 3 gametocyte genes relative to the mean levels of those genes across the 78 peripheral blood samples. **(E). Size distribution of gametocytes across organs and patients.** Size distribution for tissues with 10 or more gametocytes per 100 HPF from the initial screen. Measurements of area in square microns with made using ImageJ software, applied to Pfs16-labeled IHC images from the initial screen. 15-20 gametocytes were measured for each tissue, and are plotted with different symbols for different patients (a total of 5 bone marrow and 3 spleens across 5 patients are shown). *In vitro* gametocytes were generated and measured for comparison to known gametocyte stages (See Appendix Figure A3). Gametocytes found in the bone marrow are smaller in area, than those found in other organs.


A Immunohistochemistry-based detection of gametocytes and total parasites
Pfs16/Hematoxylin Spleen
Pfs16/Hematoxylin Bone Marrow
Pfs16/Hematoxylin Brain
pLDH/Hematoxylin Brain
Hz
hp
iRBC
cap
gm
5µm
B Gametocyte and parasite burden in organs
pLDH
Pfs16
pLDH(+) cells per 100 HPF
Pfs16(+) cells per 100 HPF
*SPLEEN
*BRAIN
HEART
GUT
*BONE MRW
LUNG
FAT
LIVER
KIDNEY
N = 6
*N = 10
C Gametocyte fraction (Pfs16/pLDH ratio) in organs
Gametocyte Fraction (%)
D Expression of gametocyte genes in organs
PF14_0748 (Gam I-II)
Pfs48/45 (Gam III-IV)
Pfs25 (Gam V)
Relative expression to blood
SPLEEN
BRAIN
HEART
GUT
BONE MRW
LUNG
LIVER
FAT
KIDNEY
N=5
E Size distribution of gametocytes in organs and patients
Gametocyte area (sq. micron)
SPLEEN
BONE MRW
IN VITRO

Sub-localization of gametocytes within the bone marrow

Localization of gametocytes inside and outside of vasculature. We aimed to further investigate sequestration of immature gametocytes in the bone marrow by expanding our survey to a total of 30 patients; specifically, to determine whether sequestering gametocytes across patients could be found associated with endothelial cells, analogous to the established adherence mechanisms of asexual parasites stages. For this purpose, we performed co-labeling of bone marrow sections with markers for the parasite (pLDH or Pfs16) and host vasculature, platelet endothelial cell adhesion molecule (PECAM-1), a marker of host endothelial vessels. pLDH-labeled ring and schizont stages could be observed in the microvasculature, with schizonts appearing in contact with endothelial cells (Figure 2.2A) as has been described previously [84]. In addition, pLDH-labeled parasites as well as Pfs16-labeled gametocytes were also found in the extravascular space of the bone marrow. Gametocytes of various stages were observed in the extravascular space, including small oat-shaped young gametocytes and larger oblong stages (Figure 2.2C). We assessed the intravascular localization of parasites and gametocytes by quantifying the number of intravascular pLDH- and Pfs16-labeled parasites in 50 consecutive labeled vessels per patient, starting at the upper left corner of each section. And we assessed extravascular localization by the number of pLDH- and Pfs16-labeled parasites observed outside of labeled vessels in 50 consecutive high power fields, from the upper left corner. We performed both assessments for each of the 26 patients. Across patients, pLDH-labeled parasites could be found within small vessels of the bone marrow, yet very low amounts of Pfs16-labeled gametocytes were found inside these vessels, giving in an overall gametocyte fraction of 3% intravascularly, across all patients (Figure 2.2B). In fact, across patients, Pfs16-labeled gametocytes were found almost exclusively externally, giving a much higher gametocyte fraction, 47%, in the extravascular space (Figure 2.2D).

Evaluation of phagocytosis of parasites vs gametocytes in the bone marrow. As a large number of hemazoin malaria pigment-containing macrophages could be observed in the bone marrow, we performed co-localization of gametocytes and total parasites with CD163, a macrophage marker, to investigate the fate of parasites within the bone marrow. We found that across patients, large numbers of pLDH-labeled parasites could be found within the cytoplasm of CD163-labeled macrophages. This occurred to a lesser extend with Pfs16(+) gametocytes, suggesting that gametocytes in the bone marrow are less susceptible to phagocytosis than asexual stages (Figure 2.2E/F). This also confirms *in vitro* data suggesting that macrophages recognize gametocytes and phagocytose gametocyte-infected erythrocytes at much lower levels than asexual-infected ones[85].

Ultrastructure of parasites in the bone marrow. Ultrastructural analysis of parasites in a patient with 45% gametocyte enrichment in the bone marrow revealed presence of a subset of erythrocytes containing large parasites that lack knob structures or any other host cell modifications (known to be critical in the process of asexual cytoadherence) (Figure 2.3D). Some of these cells could be found adjacent to erythrocyte precursor cells (Figure 2.3C), within the extravascular space of the marrow. It has recently been demonstrated that gametocytes are knobless throughout their development [33], and the morphology of the knobless parasites we observe in the bone marrow is consistent with these data. Parasitized erythrocytes exhibiting classic knob structures and other host cell modifications such as Maurer's clefts were found in a subset of parasites observed in the bone marrow (Figure 2.3A) and in high numbers in the brain (Figure 2.3E-G).

Figure 2.2. Gametocytes are present in the hematopoietic system. (A/C). Gametocyte and parasite localization in the bone marrow. pLDH and Pfs16 are detected using AP FastRed chromogen (red), endothelial cells labeled with PECAM-1 antibodies and detected with horse radish peroxidase (HRP) and DAB chromogen (brown), and nuclei are labeled with Hematoxylin stain (blue). Small ring stage (top left) and large adhering trophozoite or schizont stage (top right) parasites can be readily identified within the bone marrow sinusoids. Pfs16-labeled parasites are predominantly detected outside of the vasculature. Various shapes of developing gametocytes can be found (bottom panels). s: sinusoid. h: hematopoietic system. (**B/D). Quantification of bone marrow sub-localization for Pfs16-positive and pLDH-positive cells.** Marrows were included if 50 sinusoids could be identified using PECAM-1 labeling. This excluded 4 patients due to poor quality of preservation. For the remaining 26 samples with clearly labeled sinusoids, quantities of pLDH and Pfs16-labeled parasites are plotted for 50 sinusoids (intravascular) in B and 50 high power fields of extravascular space in C. pLDH labeled parasites could be observed both in sinusoids and in the extravascular space, while Pfs16 labelled parasites are almost exclusively detected in the extravascular space. Based on these numbers, the gametocyte fractions for intravascular and extravascular localization are shown (B/D). **(E-F). Parasite and gametocyte localization within macrophages.** Pfs16 and the constitutive parasite marker BIP are detected using AP FastRed chromogen (red), and macrophages are labeled with antibodies against CD163 and detected using TMB substrate (turquoise). While both Pfs16+ and BIP+ cells can be found in macrophages, less than a quarter of total parasites within macrophages are gametocytes, as assessed by gametocyte fraction.

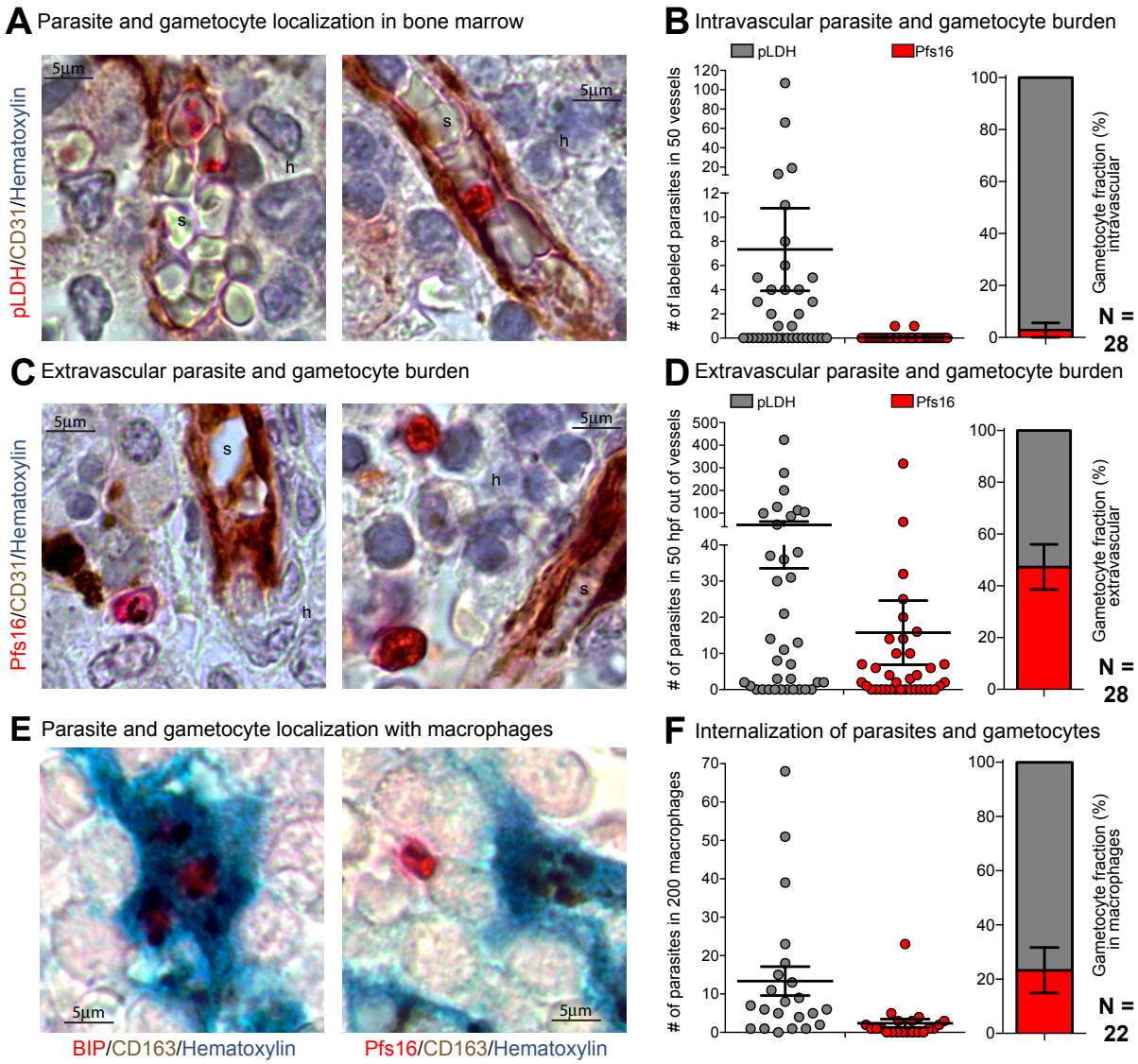
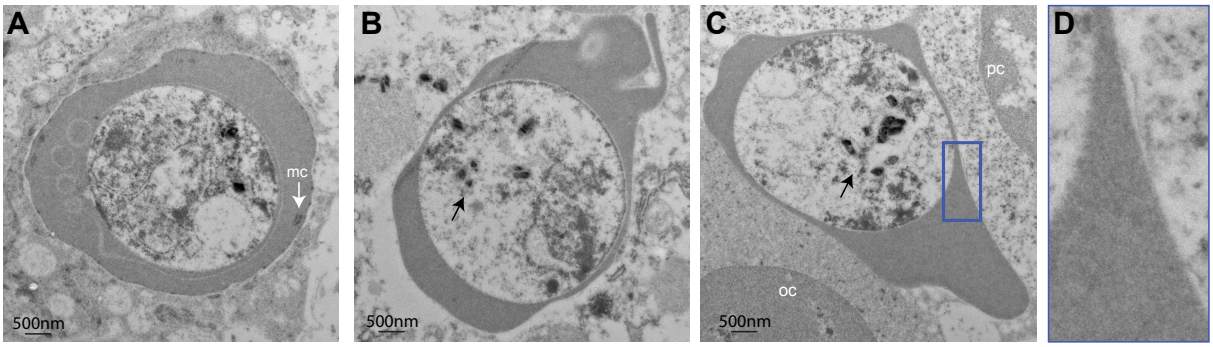
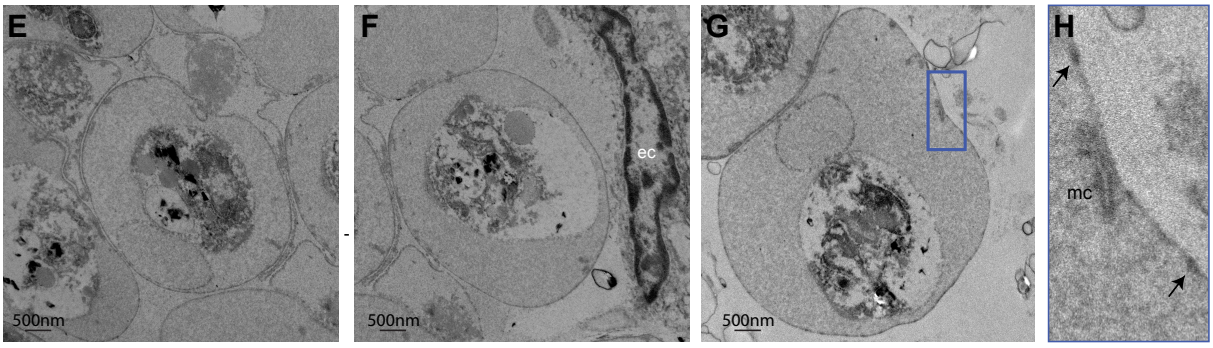

A Parasite and gametocyte localization in bone marrow
pLDH/CD31/Hematoxylin
5µm
h
s
B Intravascular parasite and gametocyte burden
pLDH
Pfs16
# of labeled parasites in 50 vessels
Gametocyte fraction (%) intravascular
N = 28
C Extravascular parasite and gametocyte burden
Pfs16/CD31/Hematoxylin
D Extravascular parasite and gametocyte burden
# of parasites in 50 hpf out of vessels
Gametocyte fraction (%) extravascular
N = 28
E Parasite and gametocyte localization with macrophages
BIP/CD163/Hematoxylin
Pfs16/CD163/Hematoxylin
F Internalization of parasites and gametocytes
# of parasites in 200 macrophages
Gametocyte fraction (%) in macrophages
N = 22

Figure 2.3. Ultrastructural features of sequestered parasites in the bone marrow. For ultrastructural analysis, we performed transmission electron microscopy on the bone marrow and brain from a patient previously observed to have a low gametocyte fraction in the brain (6%) and a high gametocyte fraction in the bone marrow (45%). (**A-D). Sequestered parasites in the bone marrow.** In the bone marrow, we observe both (A) characteristic asexual parasites (large parasite with knobs and Mauer's clefts) and (B-C) large knobless parasites characteristic of gametocytes. In C, the gametocyte is observed in close association with orthochromatic and polychromatic erythrocyte precursor cells. Apart from lack of knobs (D) other features consistent with gametocyte morphology include dispersed hemozoin (arrow, B and C) and lack of cytoplasmic modifications within the erythrocyte cytoplasm. (**E-H). Sequestered parasites in the brain.** For comparison, sequestered parasites with characteristic asexual morphology are present in brain capillaries. Parasites were observed in close contact with each other (G) and with endothelial cells (F). Knobs and Mauer's clefts can be observed on the surface of the infected erythrocyte of these cells (H). oc: orthochromatic, pc: polychromatic. mc: Mauer's clefts.

Ultrastructure of parasites in the bone marrow



Ultrastructure of parasites in the brain



Co-localization of gametocytes within erythroblastic islands.

Across patients, a large fraction of gametocytes were found to be associated with erythroblastic islands, clusters of erythroblasts that adhere to a nursing macrophage during development prior to release into the blood circulation [86]. These could be readily identified in the bone marrow as clusters of CD71-labeled cells of similar size and morphology, often appearing in contact with a hemazoin-containing macrophage (identified morphologically). We performed this assessment using Pfs16 and CD71 double-labeled on the eight patients in which at least 5 gametocytes were detected in 50 HPF. Gametocyte-infected erythrocytes could be observed in contact with erythroblast cells and/or macrophages that composed the island structures (Figure 2.4A). We quantified the amount of gametocytes in contact with erythroblast islands versus away from islands, and across patients, between 50 to 90% of the gametocytes found in the bone marrow were found in contact with these structures (mean of 72% across eight patients), (Figure 2.4B).

***In vivo* development of gametocytes within erythrocyte precursors.**

The transferrin receptor (CD71), as a marker for erythroblasts and reticulocytes, also enabled us to quantify the distribution of gametocytes inside erythrocyte precursor cells. We observed only a very small fraction of gametocytes within CD71(+) erythrocyte precursor cells and only observed this in the two individuals with the highest gametocytemia (Figure 2.4C/D). We therefore further examined the patient in which we identified the highest subset of gametocytes within CD71(+) cells and performed a size comparison between the gametocytes within CD71(+) cells and in CD71(-) cells. For this analysis, we imaged fifteen gametocytes inside CD71(+) and fifteen gametocytes inside CD71(-) cells. Using the area measurements

from our *in vitro* gametocyte stage (Figure 2.1E and Appendix Figure A3), we categorized gametocytes as Stage I or Stage II/III gametocytes (none of the gametocytes in this patient were measured in the range of Stage IV or V gametocytes). Parasites greater than 7.9 square microns in area (lower end of the range quantified for *in vitro* Stage II gametocytes) were classified as Stage II/III, and parasites under this size were categorized at Stage Is. Over 60% of the Stage I gametocytes were found within CD71(+) cells, while under 30% of the Stage II/III gametocytes were found within CD71(+) cells. This suggests that *in vivo* there is some degree of gametocyte development occurring in erythrocyte precursors, and appears to be more specific to young gametocytes than older gametocytes (Figure 2.4D).

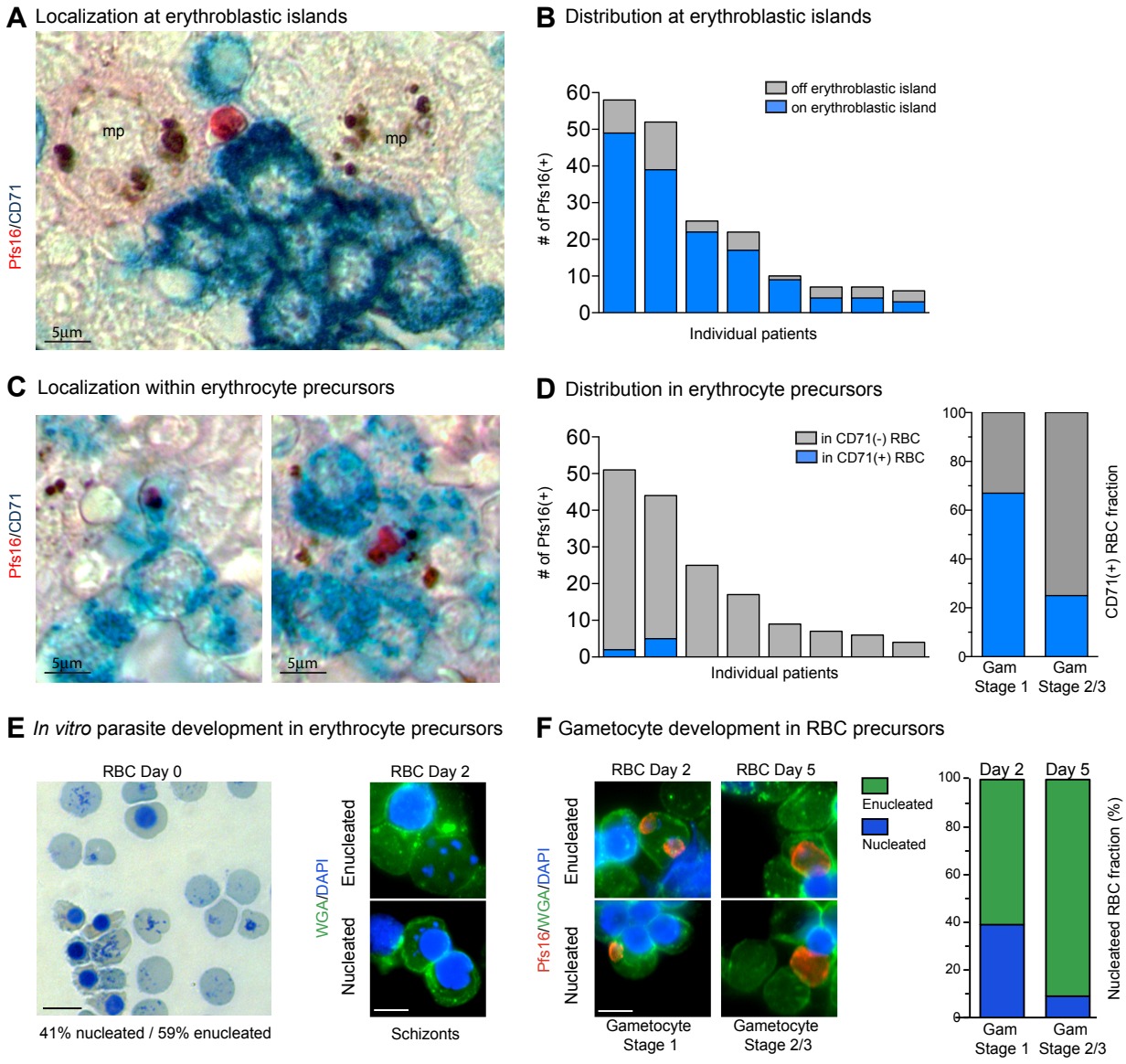
***In vitro* development of gametocytes within erythrocyte precursors.**

Given the enrichment of immature gametocytes in the extravascular space of the bone marrow and previous evidence that gametocyte formation is elevated in reticulocyte-enriched blood [62,87,88], we wanted to investigate the possibility of gametocyte formation within the hematopoietic system in a controlled *in vitro* setting. *In vitro* generated human red blood cell precursors derived from hematopoietic stem cells [89,90] were incubated with *P. falciparum* schizont stages and development of sexual and asexual progeny assessed. We observed continuous development of both asexual and sexual parasites in red blood cell precursors starting with the nucleated orthochromatic erythroblast, demonstrating that the hematopoietic system of the human bone marrow can support the entire erythrocytic parasite cycle (Figure 2.4E). At Day 0 (when parasites were added to host cells), host cell population was 41% nucleated and 59% enucleated. By Day 2, we note that asexual development occurs normally, progressing to

schizont stage (Figure 2.4E). Using Pfs16 to label gametocytes, on Day 2, young gametocytes are also visible within nucleated and enucleated cells. On Day 5, intermediate stage II to III gametocytes could be observed in these cells, suggesting that gametocyte development occurs normally within these cells. On Day 5, both gametocyte size and host cell age has increased, as indicative of higher rates of enucleation (Figure 2.4F). This suggests that the gametocyte and its erythrocyte precursor host cell can develop in parallel, and agrees with the *in vivo* data we see suggesting that early gametocytes are enriched in younger erythrocyte precursors (Figure 2.4D).

Figure 2.4. Gametocyte development within the hematopoietic system. A-D. Gametocytes associate with erythrocyte precursor cells in the bone marrow. (A-D) Pfs16 is detected using AP FastRed chromogen, and erythrocyte precursor cells are labeled with anti-CD71 (transferrin receptor) antibodies and detected with TMB substrate (turquoise). Bone marrows were analyzed for 8 patients in which at least 5 gametocytes were detected in 50 high power fields, and the number of Pfs16(+) cells is plotted for 50 high power fields (**A/B).** A large fraction of gametocytes observed in these patients were found to be are associated with erythroblast islands, specialized niches where red blood cell precursors are eventually enucleated by a nursing macrophage and released to enter the blood circulation as reticulocytes. mp: macrophage (**C/D).** Only 2 of these individuals were found to have gametocytes within CD71+ cells, and they made up a small fraction in both cases. Gametocyte size was measured using ImageJ and compared with *in vitro* sized gametocytes. For this analysis, 15 gametocytes from within CD71(+) and 15 gametocytes from within CD71(-) were imaged from the patient with the highest % of gametocytes in CD71(+) cells and plotted by proportion CD71-positivity rate for stage I versus stage II gametocytes. **E. *In vitro* parasite development in erythroid precursor cells.** Methylene blue staining was used to confirm the presence of young erythrocytes (which still contain RNA, as seen in blue label) in the starting culture, left panel. Asexual development was observed to occur within these cells. In fluorescence images, erythrocyte (wheat germ aggultinin, FITC), and nuclei (DAPI) are labeled revealing the presence of schizonts in both nucleated and enucleated cells. **F. Gametocyte development in RBC precursors.** Gametocytes (Pfs16, Rhodamine) in addition to erythrocyte and nuclei markers were used in fluorescence microscopy to visualize gametocytes in erythrocyte precursors.

Figure 2.4 (Continued). Gametocytes could be observed in both nucleated and enucleated erythroid precursor cells. Quantification of host cell is shown for each stage of gametocyte, morphologically identified based on size.


A Localization at erythroblastic islands
Pfs16/CD71
mp
mp
5µm
B Distribution at erythroblastic islands
off erythroblastic island
on erythroblastic island
# of Pfs16(+)
Individual patients
C Localization within erythrocyte precursors
Pfs16/CD71
5µm
5µm
D Distribution in erythrocyte precursors
in CD71(-) RBC
in CD71(+) RBC
# of Pfs16(+)
Individual patients
CD71(+) RBC fraction
Gam Stage 1
Gam Stage 2/3
E In vitro parasite development in erythrocyte precursors
RBC Day 0
41% nucleated / 59% enucleated
RBC Day 2
WGA/DAPI
Enucleated
Nucleated
Schizonts
F Gametocyte development in RBC precursors
RBC Day 2
RBC Day 5
Pfs16/WGA/DAPI
Enucleated
Nucleated
Gametocyte Stage 1
Gametocyte Stage 2/3
Enucleated
Nucleated
Day 2
Day 5
Nucleateed RBC fraction (%)
Gam Stage 1
Gam Stage 2/3

Discussion

The data presented here provide evidence for the existence of a genuine developmental cycle of *P. falciparum* gametocytes in the hematopoietic system of the bone marrow. We first determined that the bone marrow has a higher gametocyte fraction than any of the other organs tested, suggesting a specific enrichment process occurring at that site. We next showed that the expression of young gametocyte genes peaks in the bone marrow, while expression of mature gametocyte genes peak in a variety of organs including the bone marrow but also the spleen and heart, which could be a result of the presence of circulating blood through these organs. We also confirmed through image measurements that vast majority of gametocytes in the bone marrow are stage I-III, while the spleen contains a subset of older stage IV-V gametocytes, further supporting the notion that young gametocytes are specifically enriched in the bone marrow.

We demonstrate that developing gametocytes are present almost exclusively in the extravascular space of the bone marrow. While asexual ring and sequestering trophozoite/schizont stages are found inside the small vessels of the bone marrow (and in contact with labeled endothelial cells), we find that gametocytes make up less than 5% of intravascular parasites. The enrichment of immature gametocytes in the bone marrow appears to come almost exclusively from the extravascular space, in which we find that nearly 50% of all parasites are gametocytes labeled with Pfs16. Of these extravascular parasites that are split approximately down the middle between asexual and sexual parasites, just 22% of the parasites found inside macrophages were gametocytes, as compared with an estimated 78% being asexual. Given this skew, the fraction of parasites *successfully* sequestered (i.e. escaping immune system) in the extravascular space of the bone marrow is above 50%. This further suggests that the bone

marrow is likely a specialized site for sexual development, and not necessarily a favorable niche for parasite development generally.

At an ultrastructural level, we observed a subset of bone marrow-residing parasites that lack knobs and other host cell modifications known to be critically important to the process of cytoadherence in asexual stages. Considering that they are not efficiently recognized by macrophages (Figure 2.3E/F and [85]) nor induce host immune responses [93] it is possible that sequestered gametocytes lack immunogenic antigens on the host cell surface altogether. This supports our hypothesis that the underlying mechanism of gametocyte sequestration is fundamentally different from the established vascular endothelium cytoadherence model taken from asexual parasite biology.

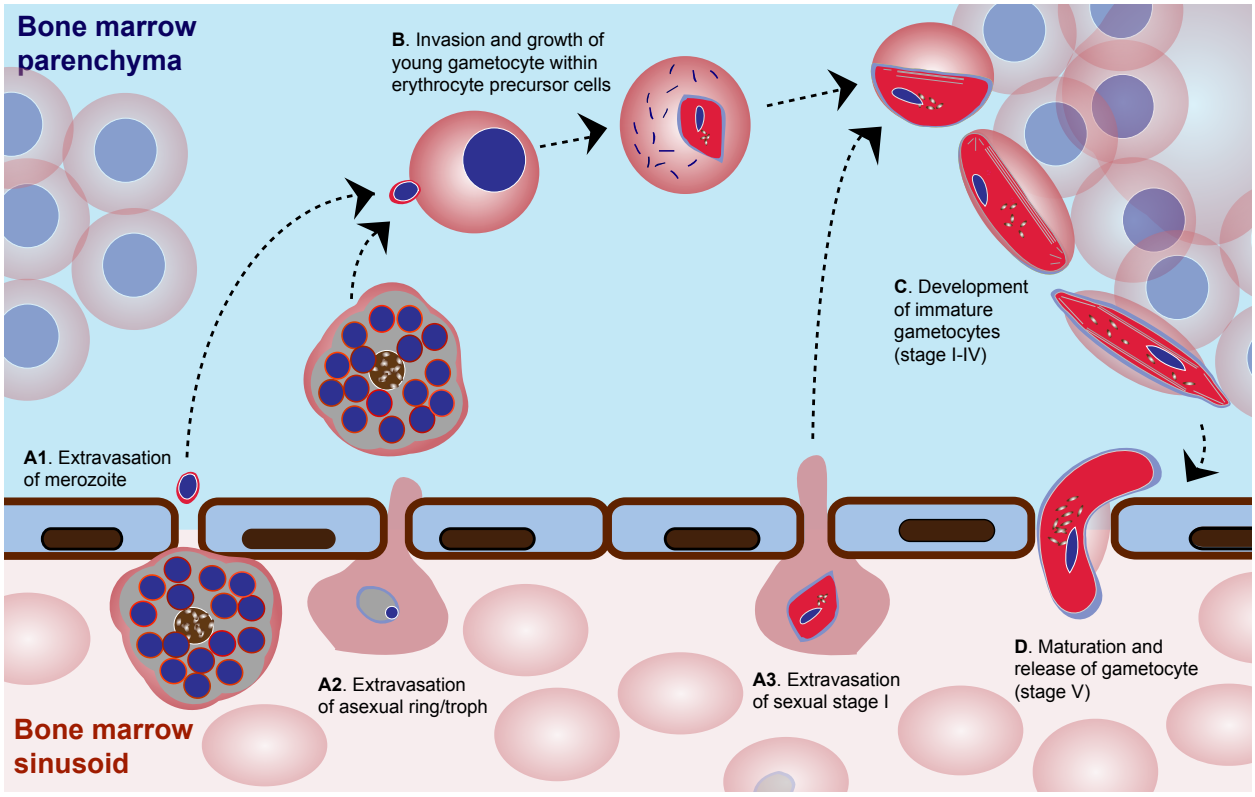
Using CD71 to label erythroblasts in the hematopoietic environment, we were able to localize gametocytes within erythroblastic island structures. Gametocytes had a very high rate of co-localization with these structures across patients, with gametocyte-infected cells appearing in close contact with CD71(+) erythroblasts and/or the nursing macrophage to which they were attached. A mean of 72% of gametocytes from eight patients were observed at these structures.

Further, we identified a small subset of bone marrow-residing gametocytes that co-localized inside CD71(+) cells. As CD71 antibody decreases significantly during erythroblast development, trailing off in the late normoblast and reticulocyte [94] and Pfs16 does not label gametocytes during the first approximately 24 hours of development (Appendix Figure A3), it is possible that the proportion of detected gametocytes within erythrocyte precursors, particularly reticulocytes, is underestimated in our analysis.

While this phenomenon was rarely observed in our *in vivo* bone marrow samples, we did observe that in the one patient for which we had enough cells to quantify, stage I gametocytes were found at a higher rate within CD71(+) cells, as compared with CD71(-) cells. Using an *in vitro* erythrocyte development system, we observed that gametocytes can grow and develop within erythrocyte precursors (orthochromatic cells and reticulocytes). This suggests that gametocyte development could occur within the hematopoietic network, which is predominantly made up of erythrocyte precursor cells. *In vitro*, we note that the fraction of stage II/III gametocytes within nucleated erythrocytes (a marker of young age) is lower than it was when these gametocytes were stage Is. Both this and the *in vivo* observation support the possibility that early gametocyte and its host cell can age in parallel.

There is a need for a paradigm shift in our conceptualization of gametocyte sequestration as an increasing body of evidence suggests that this process occurs through a process distinct from that of asexual parasites. We therefore propose a series of possible models for the enrichment of immature gametocytes in the bone marrow parenchyma.

Figure 2.5. Hypothesized models for sequestration of gametocytes in the bone marrow parenchyma. (A). Extravasation of parasites into the extravascular space of the bone marrow. Both asexual and sexual parasites are found within the extravascular space of the bone marrow (Figure 2.2). Parasites might enter as merozoite progeny from a schizont sequestered in the bone marrow sinus (A1), as asexual stage parasites, as ultrastructural bone marrow studies suggest this is possible *in vivo* [91] (A2), or as young sexual stage parasites that home to the bone marrow through receptor-ligand interactions (A3). **(B). Invasion and growth of young gametocyte within erythrocyte precursor cells.** Following schizont burst inside or just outside the marrow, merozoites will attempt to invade the nearby erythrocyte precursor cells. The young gametocyte can invade and grow within orthochromatic erythroblasts and reticulocytes (Figure 2.4). **(C). Development of immature gametocytes (stage I-IV) within the hematopoietic environment**. The gametocyte requires 8 to 10 days to develop prior to reaching maturity. During this time, the gametocyte could remain in the hematopoietic environment due to the lack of immune recognition (Figures 2.2, 2.3), and possibly with the help of a binding interaction with erythroblast islands (Figure 2.4) and/or an increase in rigidity that prevents their exit [38,39,92]. **(D). Maturation and release of gametocyte (stage V) into circulation**. The release of the mature gametocyte into the bloodstream could result from its de-attachment from the erythroblast island (possibly due to reduction in ligand expression as its host cell ages), and/or its increased deformability.


Bone marrow parenchyma
B. Invasion and growth of young gametocyte within erythrocyte precursor cells
C. Development of immature gametocytes (stage I-IV)
A1. Extravasation of merozoite
A2. Extravasation of asexual ring/troph
A3. Extravasation of sexual stage I
D. Maturation and release of gametocyte (stage V)
Bone marrow sinusoid

It remains to be shown which parasite stage (i.e., merozoite, asexual stages, or gametocyte stages) crosses the endothelial barrier to enter the bone marrow parenchyma. One possibility is that asexual schizonts, sequestering in the bone marrow sinus (Figure 2.2A and [91]) release merozoites, that cross into the extravascular compartment (Figure 2.5 A1). The small size of the merozoite stage may facilitate its passage through the discontinuous endothelial barrier of bone marrow sinusoids. Alternatively, the asexual infected erythrocyte could potentially extravasate through the endothelium (Figure 2.5 A2). In ultrastructural studies of parasites in the bone marrow, cytoplasmic processes of infected erythrocytes can be observed emerging through the endothelium and into the extravascular space [91], appearing as though these infected erythrocytes can transverse this barrier.

In either the A1 or A2 scenario, the progeny merozoites from bone-marrow-residing schizonts (either in sinus or in parenchyma) could have potentially access to the erythrocyte precursor cells of the bone marrow. Our observation that young stage I gametocytes can be found in erythrocyte precursors both *in vivo* in the extravascular space of the bone marrow as well as in our *in vitro* model demonstrates that these cells can support invasion and growth of gametocyte stages. Previous studies in *P. falciparum* and the rodent malaria model have shown increased gametocyte levels in reticulocyte-enriched blood [87,88,95]. However, it is unclear whether gametocytogenesis is triggered in the presence of these reticulocytes, if gametocyte-committed merozoites preferentially invade reticulocytes, or if gametocytes preferentially survive within these reticulocytes possibly due to their high nutrient content (Figure 2.5B). An alternative mechanism to explain this enrichment is that there is a link between schizont comittment to gametocytogenesis and the bone marrow environment. Certain environmental factors (such as the byproducts of burst erythrocytes and parasite conditioned medium) are

known to trigger gametocyte formation and therefore it is possible that the bone marrow environment itself trigger comittment to sexual development (Figure 2.5 A2). Or alternatively, that committed schizonts formed elsewhere in the body specifically home to the bone marrow due to the expression of a particular PfEMP1 ligand that result in cytoadherence in bone marrow sinusoids (Figure 2.5 A1).

Another possible pathway for gametocyte enrichment in the extravascular space is that the young gametocyte itself homes to the bone marrow through a receptor-ligand interaction, and then crosses into the parenchyma (Figure 2.5 A3). This hypothesis is supported by data suggesting that gametocytes can bind (albeit weakly) to bone marrow endothelial cells [36], and that they are still somewhat deformable in their youngest age (stage I), possibly enabling them to extravasate into the bone marrow [38,39,92]. In this model, gametocytes would be formed in the circulation, and thus, this model does not offer any link between the observed increased growth of gametocytes in reticulocyte-rich blood and the enrichment in the hematopoietic system of the bone marrow. This model instead argues that these are two independent processes, and that gametocytes would form at a higher rate only during anemic conditions when the circulating blood contains a higher proportion of reticulocytes [87,88,95].

Regardless of how the young gametocyte gets into the bone marrow parenchyma, it is unknown through what mechanism the developing gametocytes are retained within it. Adherence of erythroblasts to each other and to the central macrophage to form erythroblastic islands occurs through expression of a variety of adhesion molecules [86]. This could provide a possible mechanism for the sequestering gametocyte to bind within the extravascular space of the bone marrow, as the gametocyte would simply have to hijack a system that is already in place to retain young erythroblasts as they develop, and release them as they mature. Developing gametocytes

could adhere to these islands through their own parasite-derived ligands or possibly, if the gametocyte invaded a young erythrocyte (Figure 2.5B), the gametocyte-infected erythrocyte may still express receptors (such as erythrocyte macrophage protein, EMP [86]) involved in binding to erythroblastic islands, allowing it to stay bound inside the extravascular space attached to these structures during the host cell's development (Figure 2.5C). If the gametocyte lacks parasite-derived ligands on its cell surface (as suggested by Figure 2.2E/F and 2.3B-D), this would enable it to survive in the hematopoietic environment undetected by the immune system for the 8-10 days of development required to reach maturity. We cannot exclude the possibility that gametocytes express a ligand that is involved in adherence to such structures, yet the existing data suggest that immune system does not recognize this ligand. Alternatively, or additionally as a redundant mechanism, the increased rigidity over the course of gametocyte development [38,39,92] could lead to the retention of gametocytes within the marrow during development. Gametocytes have been shown to regain deformability upon reaching maturity [38], as well as have reduced binding rates *in vitro* to human bone marrow endothelial cells (HBMEC) in Stage V as compared to immature stages [36]. These two properties of the Stage V gametocyte could enable them to exit the bone marrow only once they have successfully matured to a deformable stage V (Figure 2.5D).

As has been demonstrated for other key processes in the parasite life cycle we hypothesize that the gametocyte may be co-opting already existing pathways of its host to sequester in the bone marrow parenchyma during its development. Our findings have wide implications for our understanding of parasite biology, and for the development and application of novel intervention strategies to block transmission of this persistent human disease.

Acknowledgements

We thank the patients' families for their consent to perform these autopsies. For assistance with the autopsies, we thank Malcolm Molyneaux, George Lioniba, Steve Kamiza, Charles Dzamala, Wales Namanya. For laboratory assistance with collection and sample handling in Malawi, we thank Jimmy Vareta and Mavis Menyere. Thanks to Dr. Maria Ericsson, Harvard Medical School Electron Microscopy Facility, for assistance on the ultrastructural analysis of tissue sections, to Dr. Michael Makler for providing pLDH antibodies, and to Evan Meyer, Andrew Frando, Abigail McFadden and Perrine Marcenac for expert technical help with IHC. Thanks to Evan Meyer and Sandra Nilsson for assistance with developing the illustration of the hypothesized models of gametocyte sequestration. This work was supported by the US National Institutes of Health (MM: R01A107755801, TT: 5R01AI034969-14). RJ was supported by a Graduate Prize Fellowship and an International Travel Fellowship from the Harvard Initiative for Global Health; SKN was supported by a fellowship from the Swedish Government.

CHAPTER THREE

Inferring Developmental Stage Composition from Gene Expression in Human Malaria

Regina Joice[1*], Vagheesh Narasimhan[2*], Jacqui Montgomery[3,4], Amar Bir Sidhu[5], Keunyoung Oh[2], Evan C. Meyer[1], Willythssa S. Pierre-Louis[2], Karl Seydel[6,11], Danny A. Milner[1,7], Kim Williamson[8], Roger Wiegand[5], Daouda Ndiaye[9], Johanna Daily[1,10], Dyann F. Wirth[1,5], Terrie Taylor[5,11], Curtis Huttenhower[2,5**], Matthias Marti[1**]

[1]Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA

[2]Department of Biostatistics, Harvard School of Public Health, Boston, MA

[3]Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Blantyre, Malawi

[4]Liverpool School of Tropical Medicine, Liverpool, UK

[5]The Broad Institute of Harvard and MIT, Cambridge, MA

[6]College of Osteopathic Medicine, Michigan State University, East Lansing, MI

[7]Department of Pathology, Brigham and Women's Hospital, Boston, MA

[8]Department of Biology, Loyola University Chicago, Chicago, IL

[9]Faculty of Medicine and Pharmacy, Cheikh Anta Diop University, Dakar, Senegal

[10]Department of Medicine, Albert Einstein College of Medicine, Bronx, NY

[11]Blantyre Malaria Project, University of Malawi College of Medicine, Blantyre, Malawi

Author contributions: R.J. and **M.M.** designed the research; **R.J., E.C.M**., and **A.B.S.** performed the research (**R.J.**: *in vitro* time course experiments, qPCR primer optimization, qPCR of patient samples, thick smear and fluoresence microscopy, **E.C.M.:** qPCR of patient samples, thick smear microscopy, **A.B.S.:** drug assays); **J.M. K.S., D.A.M., D.N., J.D.**, and **T.T.** organized and performed sample collection; **V.N., K.H. W.S.P.** and **C.H.** designed analytic tools and analyzed data. **K.W., R.W.** and **D.F.W.** contributed helpful discussion; and **R.J., V.N., C.H.** and **M.M**. wrote the manuscript.

Abstract

In the current era of malaria eradication, tools designed to detect the transmissible stage of malaria (the sexual gametocyte stage) are critical. Here, we present a method for estimating relative amounts of *Plasmodium falciparum* parasite stages from gene expression measurements. These are modeled using constrained linear regression to characterize stage-specific expression profiles within mixed-stage populations. We validated inferences both from microarrays and from quantitative RT-PCR (qRT-PCR) measurements, based on the expression of a small set of key transcriptional markers. This sufficient marker set was identified by backward selection from the whole genome as available from expression arrays targeting one sentinel marker per stage. When applied *in vivo* to peripheral blood samples from malaria patients, the system identified both immature and mature gametocyte carriers. We present this approach as a resource for staging lab and field samples alike and anticipate its applicability in epidemiological studies of gametocyte carriage.

Introduction

Gametocytes execute substantially different transcriptional programs from asexual parasite stages, as has been well studied *in vitro* [82]. Similar to the sequential dynamics of the asexual *Plasmodium* life cycle [96,97], gametocytes develop in a staged progression from young through intermediate to mature cells in preparation for meiosis and further development in the mosquito vector. Each of these stages presents a distinct expression pattern and different biology: it is hypothesized that young gametocytes are in blood circulation before the intermediate phase sequesters for 8-10 days; mature gametocytes are present in circulation for successful transmission to the mosquito vector. The switch between asexual replication and sexual gametocyte development does not occur ubiquitously *in vivo* or *in vitro*, as even the most synchronized gametocyte induction protocols result in partially asynchronous and mixed gametocyte stages [82]. This problem is compounded *in vivo*, as blood sampled during infection is likely to contain both gametocyte and asexual parasite populations, leading to a highly convolved transcriptional mixture.

In addition to the need to dissect these signatures for analysis of microarray data, it is also of interest to develop a field-applicable approach for detecting and quantifying both young (earliest indication of conversion to sexual development) and mature (indication of infectiousness to mosquito vector) gametocyte stages. Transcriptional approaches such as RT-PCR, QT-NASBA and RT-LAMP have been developed [48,49,50] using established markers of gametocyte development. Population surveys using the mature gametocyte marker *Pfs25* QT-NASBA in particular suggest that between 15-70% of cross-sectional study cohorts [48,52,53] and 39-89% of uncomplicated malaria cohorts are *Pfs25*-positive [51,54,55,56,57,58]. Surveys using the putative young gametocyte marker *Pfs16* in circulating blood were inconclusive, as

they instead revealed a quantitative relationship between *Pfs16* transcript and asexual parasitemia [49,50]. While these approaches enable sensitive detection of these transcripts, it is unclear how the detection of these transcripts - particularly *Pfs16* - relates to actual gametocyte carriage and thus transmission.

There are only a few known markers that are specific for young gametocytes [98,99], and these have not yet been successfully employed in field studies to quantify this life cycle stage. Since the peripheral blood of malaria-infected individuals contains an abundance of asexual stages, care must be taken both to identify markers that are highly expressed in young, intermediate or mature gametocytes and to computationally separate the expression of marker genes from other life cycle stages. Furthermore, the development of a quantitative assay for young and mature gametocyte stages using qRT-PCR has thus far been impeded primarily because this approach cannot easily distinguish transcript from contaminating genomic DNA in the sample when sequences are identical. Unfortunately, the majority of *P. falciparum* genes lack introns and thus have identical sequences for both RNA and DNA. Likewise, most gametocyte detection markers lack introns, and little effort has been made to expand this marker repertoire in order to design qRT-PCR assays for gametocytes. It is therefore worth identifying novel intron-containing markers for which exon-exon junction-spanning primers can be designed so that this approach can be used for *in vivo* gametocyte quantification.

Our goal was thus to develop a new transcript-based gametocyte model that addressed these challenges. Using a deconvolution approach, we identified intron-containing markers from across the full range of asexual and sexual development. In order to identify expression patterns specific to different gametocyte stages, particularly the young stages, existing *in vitro* gametocyte time point samples were re-analyzed to account for their mixed stage composition.

We further developed a qRT-PCR assay based on these results and, applying the model in reverse, established an algorithm to estimate the amounts of immature and mature sexual and asexual stages in a patient sample based on the expression of a small set of stage-specific markers—inspired by related approaches that have been used successfully in dealing with mixed cancerous/non-cancerous tissue samples [100,101] and with mixed stages of budding yeast [102].

We describe here a system for estimating the stage composition of mixed-stage samples of *P. falciparum* in laboratory and field samples. This deconvolution approach to gene expression analysis allowed us to determine the stage-specific expression of the *P. falciparum* transcriptome, to identify a small set of stage-specific markers from a time course of mixed stages, and to subsequently predict the proportions of each stage in independent samples of unknown composition. At each stage, we evaluated model error patterns using cross-validation among labeled training samples from across 5 *in vitro* developmental time courses. We finally applied the system to: i) microarray- and qRT-PCR-based data from infected blood samples from Malawian and Senegalese cohorts, and ii) microarray expression data from drug perturbations performed *in vitro*. In the field study results, we successfully predicted a higher fraction of gametocytes in those samples shown to contain gametocytes by thick smear, and show drug perturbation data as a proof of concept of a useful *in vitro* application of this tool. As a transmission-focused tool, this system can be applied in epidemiological settings, and as such will ideally support efforts directed toward reducing malaria prevalence worldwide.

Materials and Methods

Patients and Sample Collection

Malawi Cohort. Patients who enrolled in an ongoing severe malaria study [73] at the Queen Elizabeth Central Hospital during the 2009 and 2011 transmission season were included in this study. These patients were between the ages of 1 month to 14 years of age and came from Blantyre, Malawi and surrounding areas, where transmission is high and seasonal. All patients enrolled in the study met the clinical criteria for severe malaria, and severity was classified by Blantyre Coma Score [103]. The majority of patients were treated with an antimalarial drug (majority received quinine) within the 24 hours prior to being admitted in the ward. Parent or guardians of all children enrolled in the study were consented in writing in their own language by local native-speaking healthcare staff (nurse or doctor). The institutional review boards of the Harvard School of Public Health, Brigham and Women's Hospital, and University of Malawi College of Medicine approved all or parts of this study. The samples from the 2009 cohort have been described in detail in a recent publication [104]. For the 2011 cohort, a venous blood sample was drawn at admission and a 500µl sample of whole blood was added directly to Tri-Reagent BD (Molecular Research Center), mixed vigorously and stored at -80 C until processing. Simultaneously, thick and thin smears were collected and stored for later processing. Patients were classified as gametocyte-positive if they had 1 or more gametocytes in 100 thick smear high power fields (HPF). For standard gametocyte quantification by thick smear, we quantified gametocytes per 500 white blood cells (WBC).

Senegal Cohort. Patients who enrolled in a study of uncomplicated malaria in Thies, a low malaria endemicity region 70km from Dakar, Senegal in October 2008 during transmission season were included in this study. Patients who presented to Section de Lutte Antiparasitaire de Thies, with signs and symptoms of malaria were offered enrollment if they had microscopic confirmation of malaria infection and had mild symptoms. Institutional Review Board approvals were obtained from the Ethics Committee of the Ministry of Health in Senegal and the Human Subject Committee of the Harvard School of Public Health. Consent was obtained from the patient or a child's guardian. At the time of admission, a blood sample was taken from which 4.5 mL was transferred into Tri-Reagent BD, shaken vigorously for 15 seconds, and frozen at -80°C for transcriptional analyses. General demographic data, history of illness and hematocrit were recorded. Samples were transported to HSPH in a liquid nitrogen charged dry shipper, thawed in a room temperature bath and RNA isolated according to manufacturers instructions (Molecular Research Center). Simultaneously, thin smears were collected and stored for later processing.

P. falciparum asexual and sexual *in vitro* culture

A transgenic line, 164/GFP, of a gametocyte-producing clone of the 3D7 strain of *P. falciparum* was used to produce the mixed stage samples for model training and validation. This transgenic line, which aided in the quantification of gametocyte stages, produces stage-specific GFP expression under the *PF10_0164* gene promoter, as described previously [105] . A previously characterized non gametocyte-producing clone, F12, of the 3D7 strain was used to confirm stage-specificity of gametocyte markers [106]. Culture conditions were as described previously [107], maintaining the parasite line in O+ blood at 4% hematocrit in RPMI-1640 media

supplemented with 10% human serum. Cultures were kept at 37°C in a chamber containing mixed gas (5% CO_2, 5% O_2, 90% N_2). Prior to induction, asexual parasite cultures were synchronized for two cycles with 5% D-sorbitol [108], and subsequently induction of gametocytogenesis was performed according to the Fivelman protocol [74]. Briefly, asexual parasites were grown to a high parasitemia in the presence of partially spent ("conditioned") medium, and then sub-cultured at the schizont stage into new dishes containing fresh media and erythrocytes. One of two methods was used to reduce the amount of asexual stages in the cultures: Treatment with D-sorbitol was performed two days later to lyse asexual trophozoite/schizont stages and selectively enrich for unaffected early gametocytes, or N-acetyl glucosamine was added to the medium one day later and every subsequent day to selectively kill asexual stages.

Standard and Fluorescence Microscopy

In order to accurately quantify the stage distribution of parasites in our *in vitro* samples, we used a combination of standard and fluorescence microscopy. Parasite stage distribution was monitored throughout the parasite synchronization and induction protocol using Wright's Giemsa stain applied to thin blood smears. Quantification of asexual rings and trophozoite stages, as well as developing and mature sexual stages was done directly by light microscopy. In order to quantify early stages of sexual development that are morphologically similar to asexual stages, we used a combination of live imaging and immunofluorescence microscopy. Live imaging was performed using the transgenic 164/GFP line. Parasites were analyzed using the FITC channel on an inverted epifluorescence microscope (Zeiss) and quantification was done of

the proportion of GFP positive parasites (gametocytes) out of the total number of Hoechst positive parasites. Immunofluoresence assays were performed with cell monolayers on glass slides, prepared as described previously [77]. For labeling with the constitutive gametocyte marker Pfs16, slides were fixed in ice-cold methanol, blocked with 5% nonfat milk (reconstituted from dry powder), incubated with polyclonal mouse antibody against Pfs16 (1:2500) [78], washed and incubated with a secondary antibody conjugated to Alexa 488 (Invitrogen). Parasite nuclei were labeled with DAPI and quantification was done on the proportion of FITC positive parasites out of the total number of DAPI positive parasites. For time points in which we had data from both live and immunofluorescence experiments [78], the quantification of early gametocytes from both methods was averaged to give the final amount.

Primer & Probe Design

We developed a quantitative reverse-transcriptase PCR (qRT-PCR) assay for the quantification of five key gene transcripts. The assay includes primers and probes designed against the five markers: *PFE0065w* for the early asexual stage, *PF10_0020* for the late asexual stage, *PF14_0748* for the young gametocyte, *PF14_0367* for the developing gametocyte and *PF11_0209* for a constitutive marker. These are described in more detail in the Results section. Primers and probes were designed by hand using the PrimerExpress software (Applied Biosystems) and following recommended guidelines for qRT-PCR primer and probe design. Primers were specifically designed to cross exon-exon junctions, so as to reduce genomic DNA amplification. In addition both primers and probes were checked for homology against *Plasmodium* or human homologous sequences using PlasmoDB and NCBI Blast in order to

eliminate the chances of non-specific amplification (see also Appendix Table B4 for primer and probe validation).

RNA Extraction, DNAse Digest and Reverse Transcription

RNA from mixed stage cultures was preserved and extracted and processed as previously described [79]. Briefly, samples were stored in TriReagent (Molecular Research Center) until use. For sample processing, RNA was extracted by the phenol-chloroform method followed by DNAse digest (Ambion), and a second phenol-chloroform extraction for protein removal and sample concentration. For first strand synthesis we used the SuperScript III First Strand Synthesis kit (Invitrogen).

qRT-PCR Assay Optimization

Amplification of the correct target sequence was confirmed by gel electophoresis and melt curve analysis using SYBR Green (BioRad). The ability of primer pairs to discriminately amplify the cDNA product was determined by performing qRT-PCR on mixed stage *P. falciparum* cDNA and genomic DNA (same extraction, minus DNAse digest and reverse transcription for genomic DNA). The possibility of non-specific amplification with host template was ruled out by performing qRT-PCR on cDNA and genomic DNA from a human whole blood sample from which GAPDH was successfully amplified (data not shown). Primer pair efficiencies were determined by calculating the slope of the crossing threshold (CT) values on 10-fold serial dilutions of mixed stage cDNA (Appendix Table B4).

Sensitivity Assessment

Sensitivity of this assay was determined using 10-fold serial dilutions of mixed stage cDNA. Sensitivity was estimated by calculating the amount of parasites of each stage that were present in the sample at the limit of detection by qRT-PCR.

Microarray expression analysis of Senegalese patient isolates

Peripheral blood RNA (Senegal patients only) was assessed by bioanalyzer and RNA Integrity Number, and high quality RNA samples were labeled and hybridized to an oligonucleotide array (Affymetrix) custom-designed for the *P. falciparum* 3D7 genome, PlasmoFB, as published previously [97]. The raw CEL files were condensed into GCT expression files using RMA and the default parameter settings in ExpressionFileCreator in GenePattern [109].

Development and validation of the constrained linear regression model

Development. A constrained linear regression model was constructed to estimate the contribution of each stage to the overall transcript abundance from training data. Stages include: asexual ring (R), asexual trophozoite/schizont (T), immature gametocyte (IG), mature gametocyte (MG) for the 5 marker model. For the 6 marker model, IG is split into young gameotcyte (YG) and developing gametocyte (DG). For any transcriptional signature that cannot be assigned to one of these groups, it is assigned as unknown (U).

The total transcript abundance y_g for each gene g was modeled as a mixture of its abundance in

each specific stage U, R, T, IG or YG+DG and MG. The mixture fraction x_s represents the proportion of parasites in each stage s and is thus constrained between zero and one.

Microarray model training to establish genome-wide stage-specific expression. This model was first fit to estimate the contribution of each stage to each transcript's abundance, parameters $\beta_{g,s}$. Labeled training data were obtained from a published *in vitro* time course [82,97] in which the stage-specific parasite fractions x_s were known from fluorescence microscopy. The model was fit to these training data using lm, the linear model method without an intercept, in R, resulting in a table of 5,159 genes across the five erythrocyte stage-specific parameters (Appendix Table B1).

Gene Set Enrichment Analysis

GSEA was performed using the pre-ranked $\beta_{g,s}$ parameters determined in this manner in combination with a newly derived set of *P. falciparum*-specific gene sets. For this purpose *P. falciparum* Gene Ontology [110] leaf annotations were obtained from PlasmoDB and propagated into the ontology using the Sleipnir functional genomics library (34). Gene sets containing <2 genes were removed, and overlapping gene sets were merged (combining the top 10% of gene set pairs ranked by the fraction of shared genes relative to the total size of the smaller set). These gene sets were finally combined with the *P. falciparum* pathways from KEGG [111]. Enrichment of the resulting pathway sets (Appendix Table B3) was then assessed in the five pre-ranked parameter lists using GSEA over 1000 bootstraps.

Model application to predict stage distribution

Given the learned model parameters $\beta_{g,s}$ from stage-labeled data with known x_s, the model was inverted to infer the unknown stage distributions x_s in new samples. A quadratic programming approach was used to solve the system of linear equations with the constraint that the proportions of all stages must sum to 1:

$$y = \beta^{\mathrm{T}} x$$

subject to $\sum_{s \in \{U,R,T,IG,MG\}} x_s = 1$

and $x \geq 0$

We implemented this process in the R function quadprog and solved for the stage distributions using the sets of six (for microarrays) or five (for PCR data) markers ultimately selected as follows.

Co-Normalization of Reference and Inference Data Sets.

Microarray data sets. For model inference in microarray data, missing values remaining in input microarrays were imputed per dataset using the row mean across all samples. For each set of two or more training and inference datasets obtained in different batches, values used in the modeling process were co-normalized using quantile normalization. Datasets were first merged, retaining only gene names common among both datasets and the resulting merged data was jointly quantile normalized.

qRT-PCR data sets. Similarly for qRT-PCR expression values, raw values were de-logged using the formula to bring the data into a range of log-scaled abundance counts comparable to microarray samples. Missing values were imputed to the lowest detectable limit on our qRT-PCR assay, which corresponds to a CT value of 50 or 3.23 in de-logged expression space. Again considering sets of reference and inference data together, data were median normalized, such that all samples had the same median value (i.e. adjusting expression values by subtracting the column median and adding the global median).

Stepwise Backward Marker Selection

To identify reduced marker sets appropriate for inference, we used the model's inference process to perform *in silico* marker selection based on greatest accuracy of stage inference in our labeled microarray training data. Iteratively, model inference was performed on the total labeled microarray time courses using the complete transcriptome. Each individual gene was removed, the model re-applied, and the marker inducing the smallest increase in inference error removed. These steps were repeated to a minimum of five markers, and the entire process repeated 15,000 times. The resulting whole transcriptome rankings were averaged to assign a rank-average selection preference to retention of each *Plasmodium* gene as one criterion for our reduced marker set.

Bootstrap Cross Validation

The model's accuracy in each labeled microarray and qPCR training set was evaluated by 10-fold bootstrap cross validation. For each fold, one third of conditions were selected for holdout with replacement to form a test set. The model was fit as above to the remaining training data and its accuracy in predicting new expression samples was calculated using a per stage root mean square error in the test set.

Results

A regression model characterizing genome-wide stage-specific expression in *P. falciparum*

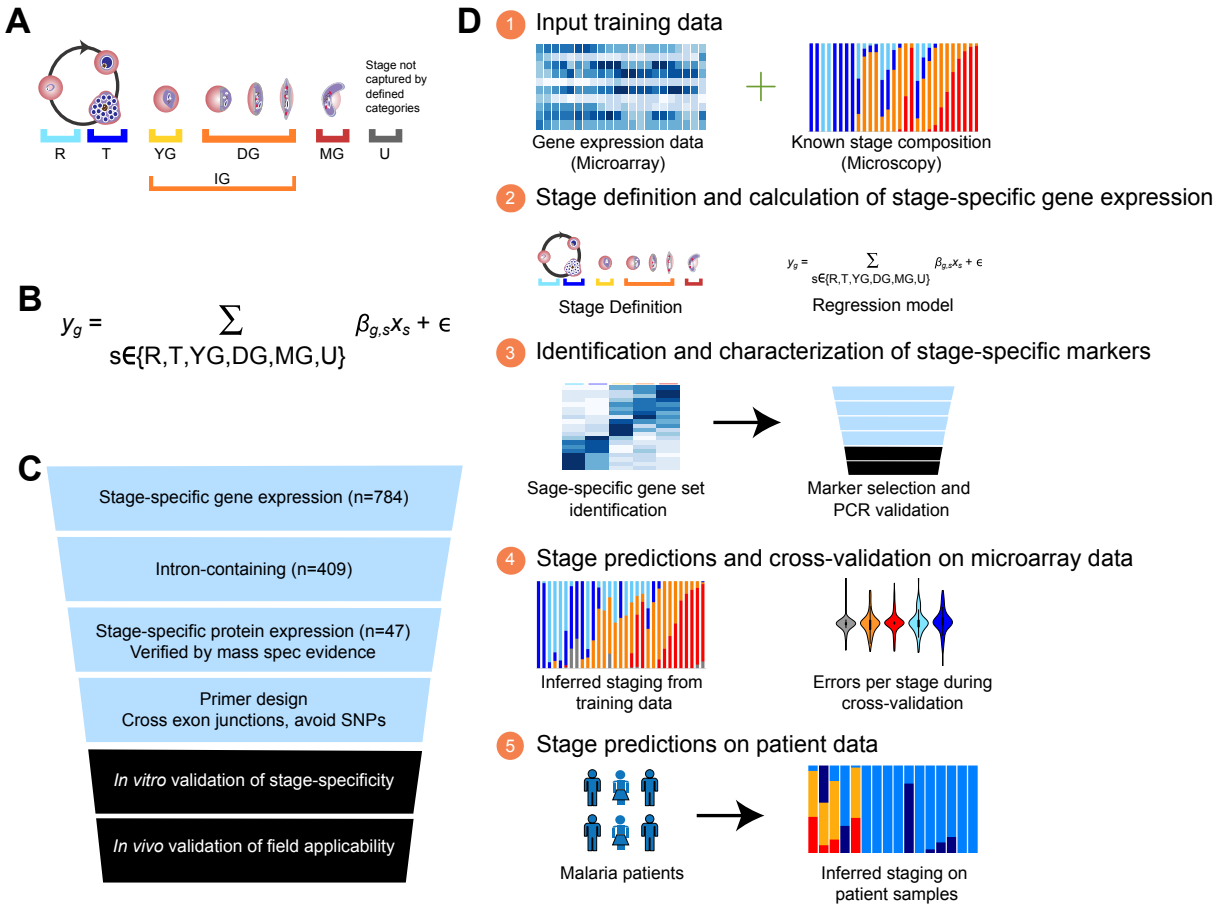
Calculation of gene-specific expression contributions to each life cycle stage. To identify expression patterns of individual stage categories while accounting for mixed stage composition, we began by using a labeled set of microarray data with asexual and gametocyte proportions assessed manually by microscopy [82,97]. For our purposes, we divided the *P. falciparum* intra-erythrocytic development into categories that reflect physiologically relevant distinctions during the course of natural infection in the human host. Specifically, we separated those phases found sequestered in tissues from those found in circulation, resulting in a total of five categories. These included two asexual categories, i) the circulating ring stage (termed "R") and ii) the sequestering trophozoite and schizont stage (termed "T"). The three gametocyte categories were i) the young gametocyte ring and Stage I (termed "YG"), which are hypothesized to circulate, ii) the sequestering developing gametocyte stages II, III, IV, (termed "DG") and iii) the circulating mature stage V gametocyte (termed "MG"). For most analyses, we grouped YG and DG together as one category encompassing all immature gametocytes (termed "IG"). We subsequently

constructed a constrained regression model to identify the degree to which each *P. falciparum* gene's expression varied with respect to life cycle stage (Figure 3.1).

The model parameters encode the relative expression of each gene that can be attributed to each life cycle stage, five of which are described above and a sixth category representing transcriptional activity not well-captured by any of these (Figure 3.1A). The model was initially fit using microarray expression data from three time courses spanning the *P. falciparum* intra-erythrocytic life cycle wherein the stage distributions at each of the time points were determined by microscopy [82,97] (Figure 3.1 D1).

In any one time point or sample, the total transcript abundance y_g for each gene g was modeled as a mixture of its abundance in each specific stage. The mixture fraction x_s represented the fraction of parasites in stage s (Figure 3.1B), and the model was constrained to require the sum of x_s across stages to remain equal to one. The contribution of each stage to g's overall transcript abundance was captured as $\beta_{g,s}$ parameters which provided not only predictive accuracy but were also used to identify stage-specific gene sets and pathways (Figure 3.1 D2). After identifying each gene g's stage-specific parameters $\beta_{g,s}$, but before winnowing them down to a minimal set of markers for qRT-PCR, we inspected the resulting genome-wide characterization of *P. falciparum* life cycle transcriptional activity in order to identify stage-specific pathways and regulatory mechanisms.

Figure 3.1. ***In silico*** **dissection approach involving a linear regression model to identify stage-specific gene expression profiles within bulk population gene expression. (A). Definition of physiologically relevant stage categories within *P. falciparum* development for which we will identify stage-specific gene signatures.** Stages are as follows: R: asexual ring, T: asexual trophozoite and schizont, YG: young gametocyte ring and stage I, DG: developing gametocyte stages II, III, and IV, IG: all immature gametocytes (YG + DG), MG: mature gametocyte stage V, and U: unexpected profile not captured by our defined stages. **(B). Linear regression model for the deconvolution of bulk gene expression data from mixed stage samples.** Terms are as follows: y_g: total expression of gene g, $\beta_{g,s}$: expression of gene g attributed to stage s, X_s: proportion of the sample that is stage s. **(C). Marker Selection.** Filters used to narrow down gene sets to our set of sentinel markers for field-applicable qRT-PCR assay. As we chose markers for ring and trophozoite/schizont stages *a priori* based on published stage-specific gene expression data for asexual development [96,97,112], we used this selection method to identify markers for the remaining gametocyte stage categories. **(D). Overall stage prediction schematic.**


A
R
T
YG
DG
MG
U
IG
Stage not captured by defined categories
B
$y_g = \sum_{s \in \{R,T,YG,DG,MG,U\}} \beta_{g,s} x_s + \epsilon$
C
Stage-specific gene expression (n=784)
Intron-containing (n=409)
Stage-specific protein expression (n=47)
Verified by mass spec evidence
Primer design
Cross exon junctions, avoid SNPs
In vitro validation of stage-specificity
In vivo validation of field applicability
D
1 Input training data
Gene expression data (Microarray)
Known stage composition (Microscopy)
2 Stage definition and calculation of stage-specific gene expression
Stage Definition
Regression model
3 Identification and characterization of stage-specific markers
Sage-specific gene set identification
Marker selection and PCR validation
4 Stage predictions and cross-validation on microarray data
Inferred staging from training data
Errors per stage during cross-validation
5 Stage predictions on patient data
Malaria patients
Inferred staging on patient samples

Definition of stage-specific gene sets reveals parasite biology. The model parameters $\beta_{g,s}$ provided a measure of the amount of expression of each gene attributable to each stage s. To initially identify genes with stage-specific regulation, we selected those genes for each stage where $\beta_{g,s}$ was at least two (for sexual stages) or one (for asexual stages) standard deviations further above mean than for any other stage. This process resulted in 637 stage-specific genes distributed across five stages: 154 (R), 34 (T), 229 (YG), 34 (DG) and 186 (MG), with the top fifteen individual stage-specific markers appearing in Figure 2A (see Appendix Table B1 for genome-wide analysis).

Gene set enrichment analysis. Prior to selecting individual stage-specific genes from these data, stage-specific pathway activity was assessed genome-wide using Gene Set Enrichment Analysis (GSEA). Since the standard gene sets available for GSEA are somewhat sparse for *P. falciparum*, a pathway database was constructed using the gene annotations in PlasmoDB for the Gene Ontology [110] (see also Appendix Tables B2 and B3). These gene sets were then used with GSEA on our model parameters, employing the genome-wide $\beta_{g,s}$ parameters z-scored across stages as a pre-ranked statistic for enrichment testing. The resulting stage-specific pathway enrichments included biological activities that are known to be associated with particular stages in the asexual cycle, such as host cell remodeling in ring stages (R) and host cell invasion and replication in the later stages of the asexual cycle (T). Our analysis supports earlier observations from the initial gametocyte transcriptome study and a more recent early gametocyte protreome [82] that mitochondrial and lipid metabolism are significantly up-regulated during early sexual development [82], as well as factors involved in cell cycle control. We also observed enrichment of endocytic pathways and cytoskeletal remodeling during later stages of

sexual development. These processes are likely linked to hemoglobin uptake and exflagellation during male gametogenesis, respectively (Figure 3.2B and Appendix Table B2).

Signatures of natural selection. Interaction with the host immune system can result in a higher rate of single nucleotide polymorphisms (SNPs) in genes involved in these interactions. Since life cycle stages vary in their level of interaction with the host immune system, we were interested in whether any of our stage-specific gene sets varied in signatures of natural selection. As a measure for balancing selection within a parasite population, we determined the genetic diversity of individual genes within the three gametocyte stages and two asexual stages (Figure 3.2C). For this purpose, we calculated SNP $\boldsymbol{\pi}$ values for each gene based on a sequence comparison from 25 culture-adapted strains from Senegal [113]. SNP $\boldsymbol{\pi}$ quantifies the average number of pairwise differences among a set of strains at a set of assayed SNPs. As a complementary measure for positive selection between populations, we also calculated Fst [114] when comparing the 25 strains from Senegal with a reference line from Honduras, HB3 [115]. Our analysis demonstrated that none of the genes from within the set of highly stage-specific asexual and sexual markers are under increased selection compared to the genomic average, neither within a population nor between two populations. This suggests that the stage-specific transcripts identified by this process are enriched, by these criteria, for gene products core to the regulatory program or stage transitions themselves, excluding proteins that would interact more directly with the host. Such conserved genes are likely to serve as robust markers in lab and field samples alike.

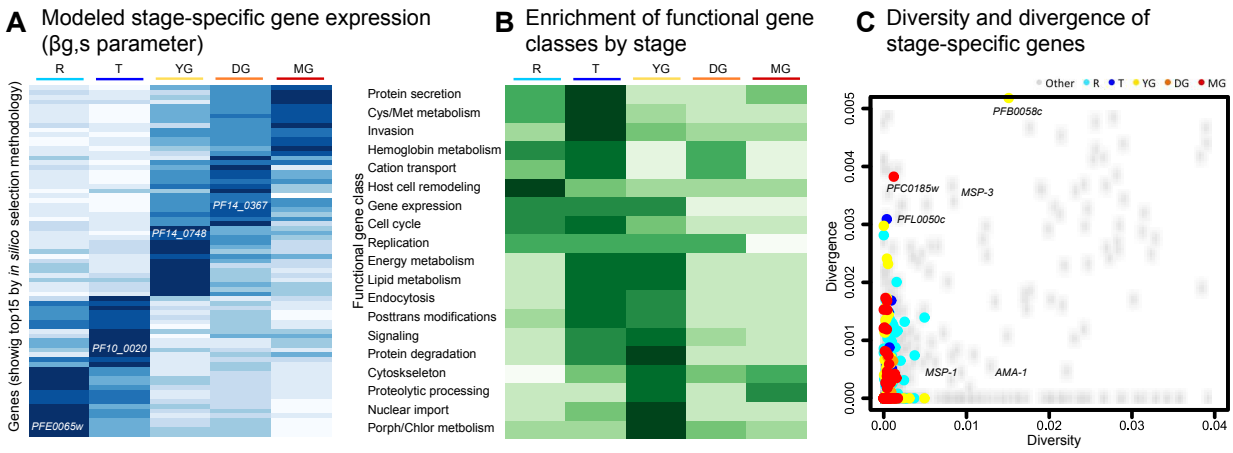


Figure 3.2. **Stage-predictive gene sets are enriched for specific biological processes but show no signature of selection by diversity/divergence measures**. **(A). Top 15 model $\beta_{g,s}$ parameters specific to each stage**; values indicate for each gene the degree of its expression attributed to each stage. **(B). Gene set enrichments of GO and KEGG processes** by stage (Appendix Table B2**). (C). Diversity (within patient) vs. divergence (between isolate)** of the *P. falciparum* genome (see methods for data sources), highlighting genes identified as stage-specific. Several known diverse/divergent (antigenically variant) markers are labeled for comparison.

Selection and validation of a sentinel marker set for stage prediction

Marker selection overview. After characterizing the genome-wide sexual stage-specific gene expression of *P. falciparum*, we proceeded to identify a small subset of markers to sufficiently recapitulate genome-wide resolution for stage prediction. To combine predictive accuracy with biological interpretability, we chose to identify the single markers representing each stage category while still minimizing prediction error. This resulted in a set of filtering criteria to obtain markers validated to be stage-specific and suitable for use in qRT-PCR (Figure 3.1C and 3.1 D3).

Selection of asexual-specific and constitutively expressed markers. To identify optimal markers for the better studied asexual R and T stages, we began with the stage-specific gene sets identified for each stage and first filtered based on presence of intron(s). We also required mass spectrometry evidence confirming the protein is expressed during asexual development, as determined using the available databases as published through the *Plasmodium* Genomics Resource Database (PlasmoDB). Next, we ranked genes based on frequency of retention during model predictive marker selection (see Methods) and used raw expression information such as lack of allelic expression variation, high absolute expression levels [97], and stage-specific expression evidence by expression timing [96] as additional filters. Based on this process, we selected *PFE0065w* for the early asexual stage (R). It encodes skeletal-binding protein 1 (SBP1), a well-characterized component of parasite-induced membrane structures in the host red blood cell (RBC) termed Maurer's clefts [112,116]. Using these same criteria, we selected *PF10_0020* as the late asexual stage (T) marker. The gene encodes a protein that is predicted to be secreted

into the host RBC [117] and contains a putative alpha/beta hydrolase domain. It is highly specific for late asexual stages not only based on the $\beta_{g,s}$ parameter, but also based on its independent gene expression profile from a comparative transcriptional analysis of three *P. falciparum* strains [96]. For the baseline marker, we chose a constitutively expressed transcript as determined by ranking all *P. falciparum* genes by the lowest standard deviation across life cycle stages [82,97] and patient isolates [80] and using a cut-off for high expression (Appendix Table B1). The selected marker *PF11_0209* encodes a conserved protein of unknown function.

Selection of gametocyte stage-specific markers. To select gametocyte markers, we began with the stage-specific gene sets identified for each stage category as above (YG, DG, or MG) and filtered first based on the presence of intron(s). Next, we retained only those genes in which mass spectrometry data again confirmed expression in the gametocyte development and absence in asexual development, providing evidence both for stage-specific activity and lack of evidence for non-specific stages. These criteria revealed a shortlist of gametocyte-specific candidates with high predictive accuracy, but in the case of the MG stage, no markers showed high and stage-specific expression levels (as defined by the $\beta_{g,s}$ parameter) comparable to the currently used gametocyte marker *Pfs25*. We therefore decided to include *Pfs25* as a MG marker to model stage distribution in microarray data, as the major criterion against using it for qRT-PCR (i.e. absence of an intron), does not apply to the analysis of chip-based data. For the remaining gametocyte stage categories in which highly expressed markers were successfully identified, we ranked gene lists based on frequency of retention during model predictive marker selection and selected *PF14_0748* and *PF14_0367* to represent young (YG) and developing (DG) gametocytes, respectively. *PF14_0748* has previously been identified as an early gametocyte marker [82,98],

and we have recently demonstrated that its promoter can drive gametocyte-specific reporter expression [105]. It encodes a protein predicted to be exported into the host RBC with a putative PHIST domain [117]. *PF14_0367* has not been described previously.

Initial validation of the model on microarray samples. With a set of six sentinel markers identified, we proceeded to test their ability to predict stage distributions *in silico*, first using cross-validation over the available labeled microarray data (Figure 3.1 D4). Having fit the complete regression model to establish $\beta_{g,s}$ parameters, we subsequently retained only those six markers' values for prediction of a sample's stage composition. This was done by solving for X_s, the percent composition of each stage, while employing a quadratic programming approach to constrain the proportions of these stages to sum to 1. Making predictions based on gene expression from our selected marker set, we successfully predict the transition from asexual to sexual development across the life cycle time course (Figure 3.3A/B). Prediction errors generated through cross-validation were minimal, with total per stage, per sample, root mean squared error of 0.195 (Figure 3.3C).

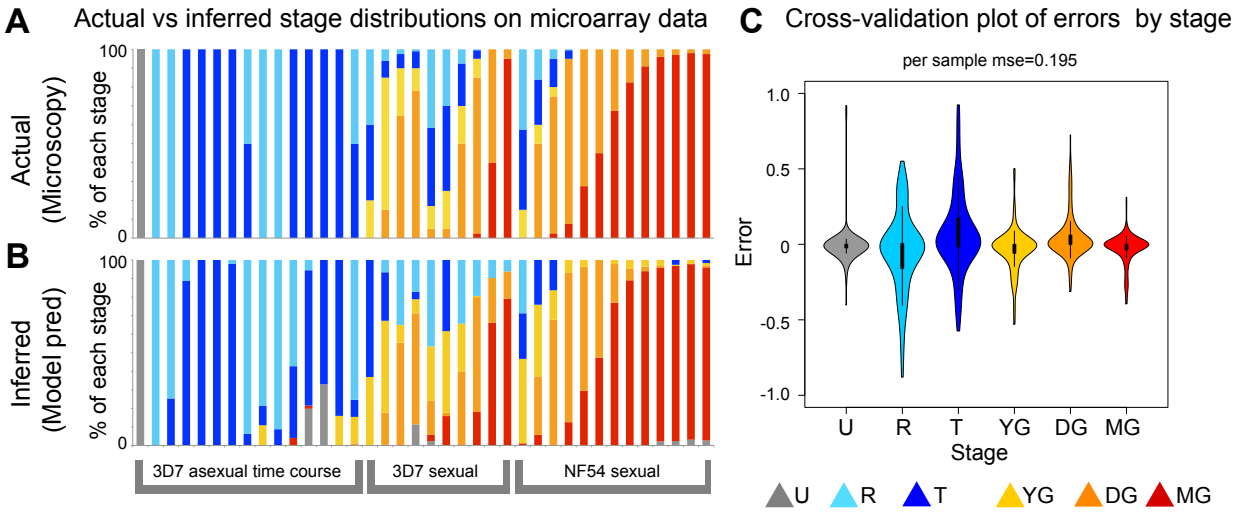


Figure 3.3. Marker selection yields a set of sentinel markers with high predictive accuracy. **(A) Actual and (B) inferred stage distributions across five microarray time courses** (two asexual and three sexual) with reference stage distributions determined by microscopy. Six markers were used to make these predictions, five as identified through filtering criteria (Table 3.1) and the previously established mature gametocyte marker *Pfs25*. **(C). Bootstrap cross-validation of error rates expected per stage in model inferences.** Violin plots show expected density, with internal boxplots detailing the 25th – 75 th percentiles and 1.5x fences.

Table 3.1. Genes used in qRT-PCR Assay.

Name	Accession	Stage-Specificity	PCR Efficiency	Limit of Detection
Skeleton-binding protein 1 (SBP1)	*PFE0065w*	ring	90.43%	10^1 pg cDNA (appx 30 rings)
Alpha-beta hydrolase, putative	*PF10_0020*	trophozoite/ schizont	87.82%	10^3 pg cDNA (appx 1000 troph/schizonts)
Plasmodium exported protein (PHISTa)	*PF14_0748*	early - mid gametocyte	92.17%	10^2 pg cDNA (appx 30 immature gametocytes)
Conserved Plasmodium protein	*PF14_0367*	mid - late gametocyte	89.77%	10^3 pg cDNA (appx 20 mature gametocytes)
Conserved Plasmodium protein	*PF11_0209*	all stages	92.06%	10^2 pg cDNA (appx 500 total parasites)

Table 3.1. Details of the qRT-PCR compatible marker set selected by our combined filtering process. PCR efficiency was calculated based on the slope of the line after running a series of 10-fold dilutions of mixed-stage cDNA (Appendix Figure B3). Limit of detection was calculated based on the number of parasite stages estimated to be present in the last dilution where the marker was detected. Detection limit ranged from $10^1 – 10^3$ pg cDNA, corresponding to approximately 20 - 200 cells, depending on the stage.

qRT-PCR gametocyte stage quantification using the sentinel markers

We next sought to test an extension of the microarray-based model to transcriptional measurements obtained by PCR, which might eventually be more appropriate for a field assay. As no MG marker that achieved our filtering criteria (see Figure 3.1C) for qRT-PCR also matched both the high expression levels and stage-specificity of the existing *Pfs25* marker for gametocyte detection, we assessed the utility of the YG and DG markers in combination for the prediction of gametocyte quantities when applying the model to qRT-PCR data. Specifically, *PF14_0748* and *PF14_0367* were likely to represent immature (IG) and mature (MG) gametocytes in combination, as *PF14_0367* had a $\beta_{g,s}$ parameter similar to that of *Pfs25* in mature gametocyte stages. We therefore cross-validated this 5-marker PCR set (Table 3.1) comparably to the 6-marker microarray set, using the *in vitro* microarray time courses as described above. The simplified model remained able to predict stage distribution accurately, with a root mean squared error comparable to that of the 6-marker model (Appendix Figure B1).

In order to create a qRT-PCR assay for our sentinel transcripts, we designed exon-exon junction spanning primers (distinguishing transcripts from genomic DNA) and sequence-specific probes (distinguishing transcripts from non-specific background amplification). Following confirmation that our primer/probe sets selectively amplified cDNA and not DNA or non-specific products, we validated the stage-specific expression using *in vitro*-derived asexual and sexual stage RNA (See Appendix Figure A3 for optimization and validation of qRT-PCR parameters). For this, we used the gametocyte-producing reference line 3D7 and a gametocyte-deficient clone thereof (termed F12 [106]) to confirm the stage-specificity of each of our sentinel markers. Normalized expression data from time courses of 3D7 and F12 confirmed stage-specificity of our sentinel marker set (Appendix Figure B3). The asexual markers alternate with

respect to time points in which there were predominately rings or trophozoites and schizonts in the culture, with similar results for both the F12 and 3D7 lines. The sexual markers demonstrate stage-specificity within the 3D7 time course and no appreciable expression in the F12 line once normalized (Appendix Figure B3). Specifically, *PF14_0748* expression is detected in the early and mid gametocyte time points, while *PF14_0367* expression is detected in both mid and late time points.

***In vitro* stage prediction using qRT-PCR.** Having validated the stage-specificity of each marker *in vitro* and tested their predictive-ability as a set *in silico*, we next compiled a larger set of *in vitro* samples in order to tune a set of model parameters for stage prediction specifically using qRT-PCR-based expression measurements. This required a set of *P. falciparum* expression samples assayed by qRT-PCR and labeled with known distributions of parasite life cycle stages. For this purpose we used the data from our *in vitro* validation on the 3D7 and F12 lines, in combination with additional *in vitro* samples generated from across a 48-hour period of asexual development, and a two-week period of gametocyte conversion. For the additional data points, we used a transgenic line in 3D7 background, termed 164/GFP, that expresses fluorescent protein under the gametocyte-specific *PF10_0164* promoter [105]. This parasite line enabled us to determine stage composition with high accuracy by both Giemsa stain and fluorescence microscopy throughout gametocyte development and starting at the earliest stages (Appendix Figure B4). The model was trained on these qRT-PCR datasets as described earlier for microarray data, again determining the contribution $\beta_{g,s}$ of each stage to our five markers' expression and performing cross-validation to evaluate the final prediction error rates per stage. Again, we see that the model accurately predicts the absence of sexual stages in F12 and low

parasitemia cultures of 3D7, while predicting young gametocytes in the high/stressed 3D7 cultures, and developing and mature gametocytes in the later mixed gametocyte induction cultures (Figure 3.4C).

Application of the qRT-PCR and model predictions to a cohort of severe malaria cases. To test the sensitivity of our system in an epidemiological context, peripheral blood samples from a cohort of severe malaria patients in Blantyre, Malawi were collected over the course of the malaria transmission season in 2011 (see Table 3.2 for patient cohort characteristics). Ring parasitemia was quantified using standard methods (thin smear microscopy and grid) and gametocyte quantification (thick smear microscopy screen of 100 high power fields for positivity, then quantification against 500 white blood cells, see Figure 3.4D). Of the 86 samples examined, 8 were gametocyte-positive by thick smear. These 8 samples were matched with 8 gametocyte-negative samples with equivalent parasitemia and qRT-PCR was performed. Seven of the 8 (88%) microscopy-positive gametocyte carriers were qRT-PCR positive for one or both gametocyte markers, while only three of the 8 (38%) matched microscopy-negative individuals were positive for either gametocyte marker, with highest levels of expression were observed in the gametocyte carriers (Figure 3.4E). The model predicted that most patients have a ring-dominated profile, as we expect to see in peripheral blood (Figure 3.4F). A number of individuals were predicted to have trophozoites and schizonts, which has been shown to be associated with severe malaria [118] and was observed in a subset of the smears from this cohort. The model predicts gametocyte fractions in four patients, all of which were gametocyte-positive by thick smear.

Figure 3.4. qRT-PCR Assay Optimization. (A). Absolute number of parasites stages that went into each qRT-PCR reaction well for a range of *in vitro* time points with varying contributions of asexual and sexual stages, from both gametocyte-producing and non-producing lines of 3D7. **(B). Relative qRT-PCR-based gene expression of stage-specific markers for T, IG and MG** are shown for time points corresponding vertically to those in part A. **(C). Inferred proportion of each stage in the total parasite load (model predictions)** are shown corresponding vertically to the time points in A and B, plotted as a percentage of total parasites in that sample. **(D). Absolute number of parasites stages per uL of blood, as determined by microscopy for *in vivo* peripheral blood samples from severe malaria patients in Blantyre, Malawi. (E). Relative qRT-PCR-based gene expression of stage-specific markers for T, IG and MG** (normalized to *SBP1*; value of *SBP1* not plotted because this asexual ring marker was detected in all samples) are shown for time points corresponding vertically to those in part D. **(F). Inferred proportion of each stage (model predictions) are shown corresponding vertically to the time points in D and E.** Stars indicate subjects in which gametocytes were observed by thick smear examination (1 or more gametocytes in 100 high power fields).

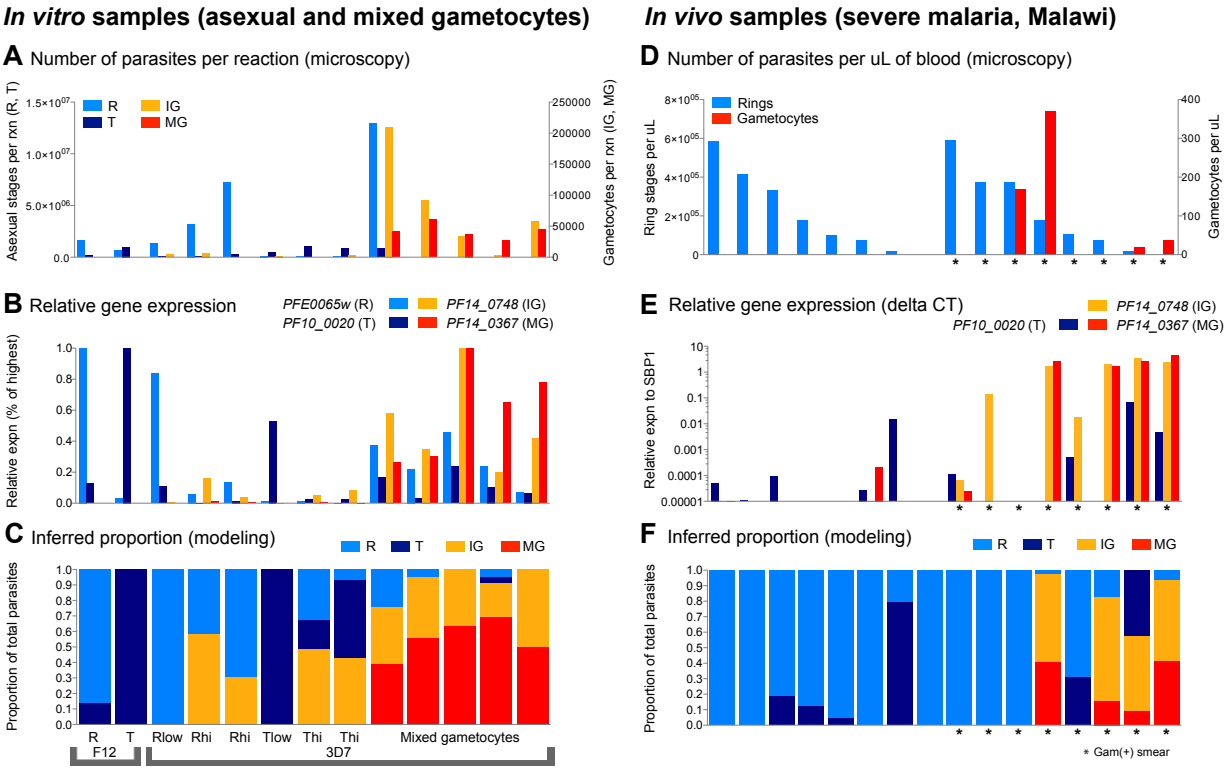

In vitro samples (asexual and mixed gametocytes)
A Number of parasites per reaction (microscopy)
Asexual stages per rxn (R, T)
Gametocytes per rxn (IG, MG)
R
T
IG
MG
B Relative gene expression
PFE0065w (R)
PF10_0020 (T)
PF14_0748 (IG)
PF14_0367 (MG)
Relative expn (% of highest)
C Inferred proportion (modeling)
Proportion of total parasites
R T RIow Rhi Rhi TIow Thi Thi Mixed gametocytes
F12
3D7
In vivo samples (severe malaria, Malawi)
D Number of parasites per uL of blood (microscopy)
Rings
Gametocytes
Ring stages per uL
Gametocytes per uL
E Relative gene expression (delta CT)
PF14_0748 (IG)
PF10_0020 (T)
PF14_0367 (MG)
Relative expn to SBP1
F Inferred proportion (modeling)
Proportion of total parasites
* Gam(+) smear

Table 3.2. Admission characteristics of severe malaria patients tested by qRT-PCR, Blantyre, Malawi, 2011.

Demographics	
Mean age (months)	52.9
Age range (months)	5-156
Gender	47.3% female
Parasitemia	
Geometric mean parasite density/µl [95% CI]	48349 [29051 – 80466]
Median parasite density/µl	74800
Parasite density range/µl	69 - 945500
Gametocytemia	
Gametocyte prevalence (Screen of 100 HPF on thick smear)	11.9% (8/67)
Gametocyte density range/µl (Counts per 500 WBC on thick smear)	0 - 370

Table 3.2. Blood was sampled from participants who met the clinical case description of cerebral malaria during the malaria transmission season in 2011. All patients were from Blantyre, Malawi and surrounding areas. Parasitemia was measured by microscopy and qRT-PCR was performed on an RNA sample stored in TriReagent BD.

Application of the expression-based gametocyte stage prediction process to *in vivo* malaria cohort and *in vitro* drug perturbation microarray samples

To gauge how this modeling process performed on patient microarray samples, in addition to the more challenging qPCR measurements, we applied it to microarray data from two patient cohorts, i) a previously published cohort of severe malaria patients from Blantyre, Malawi collected in 2009 [104] and ii) a cohort of uncomplicated malaria patients from Dakar and Thies, Senegal collected in 2008. While no staging information was available for the Senegal patients, a subset of Malawi patients were previously identified as gametocyte-positive by thick smears. The model inferred that the majority of patients from both cohorts have a strong ring-dominated profile, with the next largest subset being late asexual stages (trophozoites and schizonts) (Figure 3.5A/B). For the 10 Malawi samples in which gametocytes were observed by thick smear, our model correctly identifies 4 (40%) as such, with 0 false-positive developing or mature gametocytes predicted among the 48 thick smear-negative patients (Figure 3.5A). Interestingly, two thick smear-negative patients are predicted to have young gametocytes, which are difficult to identify by thick smear microscopy due to their morphological similarities with asexual stages.

Evaluation of gametocyte carriage in uncomplicated malaria cohort *in vivo*. A subset of the uncomplicated malaria patients from Senegal were also predicted to be gametocyte carriers (6 of mature, 1 of developing and 1 of young gametocytes). As microscopy-based information was unavailable for the Senegalese cohort, we assessed how our gametocyte inferences correlated with patient parameters. Of the 6 parameters we measured for this cohort, illness duration

differed significantly between the group of patients inferred to be gametocyte carriers and those inferred to be gametocyte-negative. The former had a longer duration of illness (6.33 days ± 1.02 SEM) than the latter (3.84 days ± 0.25 SEM, t-test p= 0.0014). This finding agrees with published data on clinical correlates of gametocyte carriage: illness duration longer than 2 days was found to be a risk factor of gametocytemia in uncomplicated malaria [62].

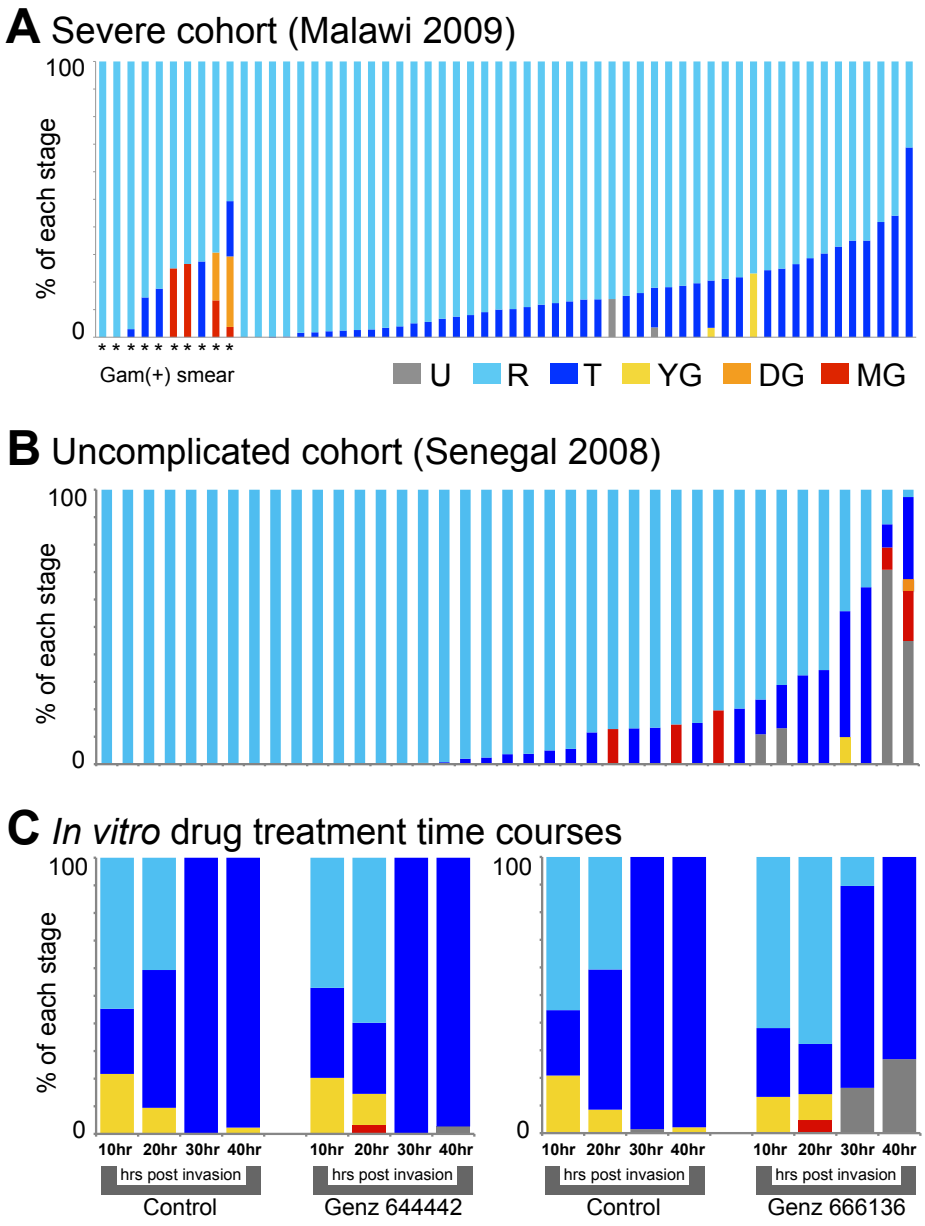
Evaluation of drug treatment on stage distribution *in vitro*. We concluded by using the model to profile the effects of drug treatments on stage distribution, of particular interest as it was recently demonstrated that artemesinin combination therapies (ACTs) have limited efficacy against mature gametocytes [42]. For this, we used microarrays from *in vitro* time course experiments with a gametocyte-producing line of *P. falciparum* (3D7) in which parasites were grown in the presence of two experimental antimalarial compounds, Genz-666136 and Genz-644442. Genz-666136 is known to inhibit the parasite's dihydroorotate dehydrogenase (DHODH) enzyme, which catalyzes the rate-limiting step in *de novo* pyrimidine biosynthesis. Inhibition of DHODH therefore inhibits biosynthesis of DNA, RNA, glycoproteins and phospholipids, all of which require pyrimidines [119]. The biological target of the other compound, Genz-644442, is unknown. *Plasmodium* sexual differentiation in response to these compounds has not been previously explored.

Based on model predictions, mature gametocytes appeared in both drug-treated time course experiments at the 20 hours post invasion time point, while in the control experiment, we observe an expected progression from ring to trophozoite stage (Figure 3.5C). In both the treated and untreated control samples, we predicted an initial, subsequently decreasing fraction of the

population to consist of young gametocytes. As these time courses were performed at high parasitemia, this distribution is in agreement with what we previously observed for 3D7 at high parasitemia (Figure 3.4C). The presence of mature gametocytes at 20 hours is likely due to a differential killing effect of the drug on asexual stages versus mature gametocytes; this would induce exactly the inferred increase in the relative proportion of gametocytes.

A more complex transcriptional response was provoked during treatment with Genz-666136, in which an increasing proportion of transcriptional signature could not be assigned to one of our stage categories (shown as "unknown", gray in Figure 3.5C). It is possible that the increase in unknown transcriptional signature corresponds to an increase in dying parasites across the drug-treated growth experiment. This proof of concept experiment demonstrates the applicability of this assay to examine the impact of various perturbations on asexual and sexual stages.

Figure 3.5. Application of Microarray Model on malaria patient cohorts and drug perturbation time courses. (**A). Model inferences for 58 pediatric severe malaria patients from Blantyre, Malawi.** Stars indicate subjects in 1 or more gametocytes were observed by thick smear examination (14 patients). **(B). Model inferences for 39 adult uncomplicated patient samples from Dakar, Senegal. (C). Model inferences for *in vitro* time course experiments** in which samples were taken at 10, 20, 30, and 40 hours post-invasion. Time courses were performed in the presence of an antimalarial compound (Genz-666136 or Genz-644442) or under normal growth conditions (Control).


A Severe cohort (Malawi 2009)
100
% of each stage
0
Gam(+) smear
U
R
T
YG
DG
MG
B Uncomplicated cohort (Senegal 2008)
100
% of each stage
0
C In vitro drug treatment time courses
100
% of each stage
0
10hr 20hr 30hr 40hr
hrs post invasion
Control
Genz 644442
Genz 666136

Discussion

Several highly sensitive single-marker molecular assays are currently used to detect *Plasmodium* gametocytes. None of these existing tools have been appropriate for detection and quantification of the full course of gametocyte development; however, due primarily to the lack of a sufficiently broad panel of stage-specific markers. Further, since malaria parasite populations exist as mixtures of the different phases of the life cycle, assays combining multiple markers require customized computational analysis methods for dealing with this complexity. We combined the development of such a bioinformatic deconvolution approach with panels of stage-specific, intron-containing markers appropriate both for microarray analysis and a newly developed qRT-PCR assay. This multi-marker platform enabled us not only to detect gametocyte carriers but primarily to infer the relative amounts of sexual and asexual stages within a sample. We provide an implementation of this platform for further development and application, particularly for refinement in field settings. This process can also be adapted bioinformatically by the exclusion or inclusion of markers to answer specific questions, such as determination of sex ratios.

Our deconvolution model provided the opportunity to define stage-specific gene sets and to characterize the biology of these stages' expression programs using tools such as GSEA, even in the absence of transcriptional data from pure stage populations. For example, our GSEA analysis confirms earlier studies that suggested increased mitochondrial and apicoplast activity during gametocyte development [82,120]. Interestingly, the analysis also suggests significant enrichment of several markers related to endocytic trafficking in late gametocyte development but not in any other parasite stage. The biological significance of this observation remains to be determined. To put such findings into context and ultimately describe the gametocyte

transcriptome at high resolution, a systematic transcriptional re-analysis of the entire *P. falciparum* gametocyte cycle using isolated and synchronous gametocyte stages will be required.

Transcriptional approaches have significantly increased the sensitivity of gametocyte detection in field-compatible assays [49,50,51,57,121]. However, these have been limited to either (i) qualitative assessments of multiple gametocyte markers, i.e. RT-PCR of young and mature gametocyte markers [57], or (ii) quantitative assessments of mature gametocytes only, i.e. QT-NASBA of the gamete surface antigen *Pfs25* [49]. In order to properly define the reservoir of parasite and gametocyte carriers in the field, it is imperative to determine both the absolute parasite burden and the stage composition of parasites in the blood circulation. Challenges have prevented the development of a diagnostic that can measure the latter, such as (i) the lack of known immature gametocyte markers, and (ii) the lack of transcriptional analysis methods to identify gametocytes with high specificity in a sample containing a mixture of stages, and (iii) the lack of known intron-containing qRT-PCR compatible gametocyte markers. We therefore tackled these challenges by developing a model specific to the quantification process and ensuring that it was compatible with both microarray and qRT-PCR measurements. This is distinct, of course, from models that would focus only on sensitivity and specificity of gametocyte *detection* from such data, which represent a potentially fruitful course of future computational investigation. Instead, by incorporating relative expression values of the markers, the model allowed us to both identify a subset of patients as gametocyte carriers and additionally quantify sub-categories of immature, developing and mature gametocyte fractions within the mixture of stages in bloodstream.

Following validation of our model on samples for which stage composition was known, we applied our model to two microarray data sets in which stage composition was unknown: i) a

cohort of uncomplicated malaria patients, and ii) two *in vitro* growth experiments in the presence of drug. In the former, we found that the mean illness duration of inferred gametocyte carriers was significantly higher than that of non-carriers, in agreement with published data demonstrating that long illness duration is linked to gametocyte carriage. In the latter, we observed enrichment of mature gametocytes as well as unexplained transcriptional signature upon the addition of drug treatment to parasites. These applications demonstrate the range of potential uses for this inference tool.

Both young asexual and sexual mature gametocyte proportions vary over the duration of a malaria infection, due to different dynamics with respect to growth and clearance rates following antimalarial drug treatment [43,55]. By applying our assay to larger cohorts, we anticipate being able to define additional features of the parasite, host or environment that influence gametocyte carriage. To be practical *in vivo*, this will require further refinement of the qRT-PCR, field-applicable component of our model. As the exon-exon junction spanning primer/probe sets for 5 markers designed here represent the first attempt at a multi-marker gametocyte-staged qRT-PCR assay, further modeling of PCR-specific measurement error and careful standardization of experimental protocol for this difficult task will both improve field inferences. Like the microarray expression model, this model successfully recapitulated the transition from asexual to sexual development across multiple *in vitro* experiments even on first application. When used initially *in vivo* for blood samples from a cohort of children with severe malaria in Malawi, the system successfully identified a subset of patients as immature and/or mature gametocyte carriers. Because immature gametocytes in particular are present in the body approximately one week earlier than the more mature forms, our approach for detecting it could be used in further investigations into factors that influence gametocyte conversion *in vivo*. We

hope that such investigations will be possible in order to further understand and, eventually, impede the mechanisms of malaria transmission.

The assay and algorithm framework presented here has potential for use in epidemiological studies such as those of asymptomatic carriers, who likely represent a major reservoir for malaria transmission. Multiple such studies are already ongoing and will yield additional samples to further optimize computational models of gametocyte differentiation. This is also true of data generated from other sensitive expression platforms such as glass-slide arrays or Nanostring. The inference process may thus have applications in better understanding the natural progression of malaria in the human host, by identifying gametocytes earlier in the course of infection and determining the impact of specific drug treatments on gametocyte development. By scaling to future population-level screens, the resulting information will help better characterize the epidemiology of gametocytemia based on malaria transmission intensity, geography, climate and season.

Acknowledgments

We would like to thank the patients and their families in Malawi and Sengal for their participation in this project. We would like to thank the team of clinicians in Malawi and Senegal who admit and care for the patients, as well as the lab technicians who collected blood samples and performed routine microscopy. We also like to acknowledge Jimmy Vareta and Mavis Menyere for excellent technical support on site. This work was supported by the US National Institutes of Health (CH, 1R01HG005969; MM, R01A107755801; TT 5R01AI034969-14) the

National Science Foundation (CH: DBI-1053486), the Harvard Institute of Global Health (RJ: International Travel Fellowship Award).

CHAPTER FOUR: CONCLUSION

4.1 Gametocyte Sequestration: Old and New Hypotheses

4.1.1 Summary of Findings: Sequestration of Gametocytes in the Bone Marrow

Understanding the basic biology of gametocyte development in the human host can lead to new drug targets for processes that are crucial to gametocyte survival in the human host. In order to characterize sequestration sites of gametocytes during their development in the human host, we conducted a survey across multiple organs using post-mortem autopsy tissue from cases of cerebral malaria. We observed that across patients, immature gametocytes were found enriched in the hematopoietic system in the bone marrow, and lacked morphological characteristics of sequestering asexual stages. These findings have shed light on a previously uncharacterized process in the developmental process of the most deadly parasitic disease of humans and suggest that gametocytes sequester through a distinct mechanism to that of asexual stages. These findings will open up new avenues in malaria research to determine the exact mechanism through which this hematopoietic enrichment occurs, and explore methods to target this mechanism for therapeutic strategies.

4.1.2 Gametocyte Localization in Bone Marrow

There has been a long-standing hypothesis in the field that gametocytes sequester in the bone marrow and the spleen, based on autopsy studies showing an enrichment of immature gametocytes at these two sites [9,23]. While these studies were performed in the late 19th and early 20th century and lacked the molecular tools or large number of patients used in our current study, we confirm their earlier findings, as we also find the bone marrow and spleen to have the

highest proportions of gametocytes. What we add to that early description is new information about the localization *within* the bone marrow, which is a particularly novel finding, as these early studies were unable to describe the localization of these stages within the bone marrow or the spleen. To date, only one recent case report has demonstrated the presence of gametocytes in the extravascular space of the bone marrow [26]. In fact, most research on gametocyte sequestration in the past twenty years has focused on identifying receptors for gametocyte binding on the surface of endothelial cells, as it was hypothesized that gametocytes bind within vessels similar to asexual stages. Our work supports the notion that gametocytes develop exclusively outside of the vasculature within the hematopoietic network of the bone marrow, and does not support a model of intravascular binding for developing gametocytes.

4.1.3 Gametocyte Preference for Erythrocyte Precursor Cells

Another hypothesis that we examined in our work was the hypothesis of whether gametocytes can invade and develop within young erythrocyte precursors. There is evidence to suggest that gametocytes may develop at higher rates in reticulocytes and young erythrocytes than they do in older red cells. Higher gametocytemias have been reported in patients with high reticulocyte counts [122], in mice infected following drug-induced hyper-reticulocyte production [88] and in reticulocyte-enriched blood from anemic patients *in vitro* [87]. These studies however did not show direct evidence for the development of gametocytes within these young erythrocytes. We show that gametocytes can grow and develop within nucleated erythrocyte precursors *in vitro* and find a subset of gametocytes in erythrocyte precursors *in vivo* in the bone marrow extravascular environment.

4.1.4 Gametocyte Binding to Endothelial Cells

The binding hypothesis is one of the most rigorously examined parts of gametocyte sequestration, with a number of studies from different groups exploring the ability of gametocytes of various ages to bind to various receptors on endothelial cells [30,31,33,36]. The most rigorous of these studies, and the one that showed the highest levels of gametocyte binding, was performed using bone marrow endothelial cells [36]. Unlike the studies showing little to no binding of gametocytes to asexual receptors CD36 and ICAM-1, this study demonstrated that binding of stage III-IV gametocytes occurs to bone marrow endothelial cells and that this interaction can be disrupted by incubation with a number of antibodies, thereby identifying putative receptors for gametocyte binding. Importantly they observed that CD36 antibodies blocked asexual binding but not gametocyte binding, that ICAM-1 antibodies partially blocked binding of both asexuals and sexual stages, and that 3 novel antibodies blocked only sexual binding and not asexual binding. The three putative receptors identified through this screen of 47 antibodies were (i) CD49c or VLA-3 ($\alpha 3\beta 1$), a $\beta 1$ integrin that is involved in cell-cell and cell-matrix adhesion interactions in a range of tissues, (ii) CD166 or activated-leukocyte cell adhesion molecule, thought to be involved in adhesion of hematopoietic progenitor cells and found on bone marrow stroma, and (iii) CD164 or MGC-24, an adhesive glycoprotein found on hematopoietic progenitor cells and stroma. Interestingly, gametocyte binding was not blocked by the addition of any one antibody, but reduced by each of these, suggesting a role of multiple receptor-ligand interactions or non-specific steric hindrance in gametoycte binding [36].

It is important to revisit these observations in light of our recent characterization of gametocyte enrichment in the extravascular space of the bone marrow, often localized at erythroblastic island structures. Erythroblastic islands are the sites within the bone marrow where erythroblasts spend their development. They are characterized by a number of

erythroblasts (estimated between 5 and 30), of similar age, surrounding a central macrophage. These islands are believed to migrate toward sinusoids as the erythroblasts mature. Islands are maintained through a variety of adhesive interactions between erythrocytes and the nursing macrophage, and between erythrocytes. Three pairs of receptor-ligand pairs have been identified in these adhesion interactions: (i) erythroblast macrophage protein (EMP) is found on both the erythroblast and macrophage and binds itself, (ii) $\alpha 4\beta 1$ integrin (VLA-4) on erythroblast and VCAM-1 on macrophage, and (iii) ICAM-4 on the erythroblasts and $\alpha 5$ integrin on the macrophage. Additionally, several macrophage proteins are known to be receptors (though their erythroblast partners are unknown). These include CD69 (sialoadhesin) and CD163. It is unknown how many additional uncharacterized adhesion interactions occur between these cells, but it is believed that additional receptors may be involved [86].

Our work demonstrates that binding to erythroblastic islands could be a possible adherence mechanism for the immature gametocyte. As neither a complete set of receptors has been identified for either gametocyte binding, nor erythroblastic island composition, it remains possible that there is some overlap between the receptors involved in these two processes. For example, $\beta 1$ integrins are involved in both processes ($\alpha 3\beta 1$ involved in binding of gametocyte-infected erythrocyte, and $\alpha 4\beta 1$involved in binding of erythroblasts to macrophages).

These early binding studies should be revisited using erythroblasts and macrophages, to test binding of gametocytes *in vitro*. This could be achieved by performing rosetting assays with these cell types and gametocytes, and/or reconstituting erythroblastic islands *in vitro* and testing gametocyte adherence. This may be a more representative system to study binding interactions

of immature gametocytes inside the bone marrow parenchyma instead of the previously used bone marrow endothelial cell lines.

4.1.5 Gametocytes Detection By Immune System (Or Lack Thereof)

It is possible that the immature gametocyte may go completely unrecognized by the host immune system during its development. Whereas parasite-derived ligands localize to the surface of the asexual infected erythrocyte in order to mediate adherence and, in so doing, elicits an immune response, the surface of immature gametocyte infected erythrocytes appears to lack parasite-derived adhesive molecules, and appears to not elicit any immune response. Antibodies to KAHRP and PfEMP1 can be found in the serum of individuals who have been exposed to malaria, and phagocytosis of asexual-infected erythrocytes is an important mechanism of the host to control asexual parasitemia. Gametocytes have been shown to lack KAHRP and knob structures on their surface *in vitro* [33], and our data confirms the presence of knobless gametocytes *in vivo*. Further, PfEMP1 was shown to localize within the gametocyte, rather than on the surface of the infected cell as it does in asexual-infected erythrocytes [33]. In fact, the transport machinery required to export these parasite-derived proteins to the surface of the infected erythrocyte has not been described in gametocytes.

Immunological studies also suggest the lack of parasite-derived antigens on the surface of immature gametocyte-infected erythrocytes. In a study of immune recognition of parasitized erythrocytes, using plasma samples from 194 Gambian children, many of whom were shown to be gametocyte carriers, not a single sample recognized the surface of an immature gametocyte-infected erythrocyte. This was despite mature gametocyte- and schizont-infected erythrocytes being recognized by a subset of plasma samples [93]. Further, in phagocytosis assays using

monocytes/macropahges and polymorphonuclear neutrophils, gametocyte-infected erythrocytes were phagocytosed at very low rates compared with schizont-infected erythrocytes [85]. In our study, we also noted that phagocytosis of parasites within the bone marrow affected asexual parasites to a higher degree than it did gametocytes.

These lines of evidence suggest that the immature gametocyte infected erythrocyte is an immunologically unrecognized cell that lacks any parasite-derived ligands on the surface of the host erythrocyte. Since the gametocyte does appear to have some level of binding capacity, this raises the possibility that the gametocyte hijacks some host process such as the natural binding interactions that occur within erythroblastic islands. If in fact gametocytes have a preference for growth within young erythrocytes, it may be that gametocytes are found exclusively within a subset of erythrocytes that still express adhesive surface markers found in erythroid development (such as erythroblast macrophage protein, EMP [86]). It remains to be explored whether the gametocyte infected erythrocytes bind via host receptors.

4.1.6 Gametocyte Localization in the Spleen

The spleen has been implicated as a possible gametocyte sequestration site based on histological work from early autopsy studies [9,23], and more recently through the deformability studies of immature gametocytes [38,39,123]. The deformability studies suggest that immature gametocyte-infected erythrocytes are unable to cross through sinusoidal slits. These slits are designed to filter any defective erythrocytes and thus any rigid parasitized erythrocyte (asexual schizont and immature gametocyte-infected cells) would be filtered through the spleen's normal filtration system in the splenic sinuses. Therefore the hypothesis exists that immature gametocytes bind to endothelial cell linings of the splenic sinuses to avoid getting filtered (which

would result in their destruction). However, the lack of binding of gametocytes at high rates to endothelial cells would suggest that this may be unlikely.

There are two other possibilities for the observation of gametocytes in the spleen. First, the gametocytes found within the spleen could be enriched there due to the large amount of circulating blood found in the spleen and the enrichment of mature gametocytes in the circulation. Data from our study does in fact suggest that there is an enrichment of mature gametocyte transcript and larger size gametocytes in the spleen. However, it has been noted in the early autopsy studies that immature gametocytes are found in this organ as well (and we also note a large size distribution in the spleen in our study). Therefore, another possible explanation is that these immature stages failed to sequester in the bone marrow and thus ended up in the spleen, not sequestered, but trapped by the filtration system. A piece of evidence in support of this theory is a case report of a splenectomized individual who had immature gametocytes in addition to trophozoites and schizonts present in the circulating bloodstream [124]. Presumably, in a spleen-intact individual, the combination of both cytoadherence (parasite succeeding in avoiding immune clearance) and splenic filtration (the fate avoided by successfully cytoadhering), will result in the absence of immature gametocytes and asexual schizonts from the bloodstream. And the failure rate of cytoadherence is what is observed in the blood of a splenectomized patient. It is not well understood how successful cytoadherence is, i.e. what rate of failure we would expect to find in the blood vs. in sequestration, but the large numbers of trophozoites and immature gametocytes identified in the splenectomized individual suggest that splenic filtration of both of these stages occurs at a high rate during infection. In this case, enrichment of immature gametocytes in the spleen could merely be due to the transient

appearance of filtered immature gametocytes prior to their destruction and does not necessarily equate to this organ being a true enrichment site.

4.1.7 Gametocytes localization in other tissues

Our tissue survey revealed the presence of gametocytes in various tissues of the body. While the gametocyte percentage was typically low in all organs other than the bone marrow, their presence in organs like the brain and gut revealed that gametocytes could form in other sites of the human host. In the blood of anemic patients, higher than normal percentages of reticulocytes are found in the circulation. And if anemia results in body-wide environmental stress cues that can induce commitment to sexual development, then it is possible that gametocytes could form in other sites than the bone marrow. However, the fate of these gametocytes is unclear. These gametocytes are likely to be present in the circulating bloodstream for a short time before they end up in the spleen, filtered due to their rigidity. They may only survive if they get stuck in vessels occluded by asexual schizonts (as may be the case in the patients in our study, particularly in the organs of high vascular sequestration such as the brain), but in that case, it is not clear whether the mature forms can escape and get into the flowing vasculature.

4.2 Gametocyte Detection: Old and New Methods

4.2.1 Summary of Method: Inference of stage proportions based on gene expression data

We have developed a method for identifying stage-specific markers and inferring developmental stage composition based on gene expression. This has allowed us to successfully infer the induction of gametocytes in time course experiments in which parasites were put under stress, as well as for the identification of carriers of young and mature gametocytes from cohorts of malaria patients. This framework enabled us to dissect stage-specific transcriptional signatures out of mixed stage samples (containing young and mature asexual and sexual stages), which are found during the course of malaria infection.

4.2.2 Application of inference tool for the identification of risk factors for gametocytogenesis

It has been proposed that early gametocytes may circulate in the peripheral blood for a short time prior to their sequestration. This hypothesis is supported by data showing that infected blood containing ring stages when grown *ex vivo* can produce stage II-III gametocytes in three days, indicating that the gametocytes were already present in the blood as very young stage I gametocytes or gameto-rings [25]. While our gametocyte sequestration study suggests the formation of gametocytes could occur inside the bone marrow within erythrocyte precursor cells, we cannot exclude the possibility that gametocytes form in other locations in the body in which a gametocyte-committed merozoite comes into contact with an erythrocyte precursor. While this is most likely to occur in the bone marrow as the marrow is enriched with erythrocyte precursors, in cases of malaria anemia, higher percentages of reticulocytes are found in the circulating blood

than in healthy conditions (normally reticulocytes make up 1% of circulating erythrocytes). Thus the possibility still exists, particularly in anemic patients, that some subset of young gametocytes form within circulating reticulocytes and initially circulate in the blood stream prior to getting filtered out by the spleen.

Detection of the these young gametocyte in the peripheral blood is very difficult due to its morphological appearance that is indistinguishable from asexual stages. Molecular markers are therefore required to identify these stages. qT-NASBA using Pfs16, a constitutive marker of gametocytes, was initially used in an attempt to quantify the subset of ring parasites that were young gametocytes, but due to Pfs16's background levels of expression in asexual stages, this marker was not successful in achieving this goal.

Detecting young gametocytes in the blood stream could be potentially useful as a tool for identifying risk factors that trigger gametocyte production. Current methods for assessing risk factors for gametocyte carriage involve comparing clinical or environmental factors to the presence of mature gametocytes in the peripheral blood. However, as mature gametocytes require 8-10 days to develop, the mature stage is not an ideal marker for determining what initially led to conversion toward sexual development. The detection of young gametocyte could enable better associations to be made between risk factors and increased commitment to gametocytogenesis.

Our qRT-PCR assay and model enabled us to infer the proportion of *in vitro* and *in vivo* samples that is composed of young gametocytes, and therefore this tool represents the opportunity to identify carriers of young gametocytes in patient cohorts. The tool could be used

in parallel to study gametocyte conversion in the laboratory, to assess a shift toward sexual conversion following *in vitro* perturbations thought to induce gametocytogenesis.

4.2.3 Application of inference tool for the assessment of infection status

In addition to inferring proportions of young and mature gametocyte stages in a sample, the model also enabled us to look at young and mature asexual stages and undetermined fractions. These too have applicability in field settings.

In a typical malaria infection, rings and mature gametocytes are the only stages present in the peripheral blood stream. However, under unusual circumstances, trophozoites and schizonts can be observed in the peripheral blood, i.e. directly following drug treatment, or in cases of severe infection [118]. One possible application of the prediction of ring and schizont proportions could be for the measurement of blood schizontaemia, a sign of severe disease, and thus could be potentially useful for clinical reasons.

During an *in vitro* time course in which parasites were grown in the presence of drug, we noted an increase in the proportion of unexpected signal that could not be assigned to a known stage. While we have yet to fully understand the reasons for this observation, we recognize that the prediction of large fractions of unexpected signal could be useful in examining perturbation data. Certain antimalarials are known to slowly kill parasites, or even induce a dormant stage in which parasites are alive but remain stuck in the ring stage and do not progress normally through development until release of drug pressure. A careful study of the stage distributions produced by a panel of antimalarial drug perturbations could for instance provide information on what the 'fingerprint' of a cytostatic vs cytotoxic drug looks like. We could hypothesize for example that the stage distribution would halt in place for a cytostatic drug, versus show increased levels of

unexplained signal for a cytotoxic drug. The applicability of this tool for understanding the affect of antimalarial drugs on stage dynamics has yet to be fully explored, but this tool could potentially be useful in either in vitro or clinical settings to track the efficacy of particular drug treatments.

Further, some drugs have been hypothesized to induce the formation of gametocytes, while others have been hypothesized to selectively kill asexuals while leaving gametocytes unaffected. The stage inference model could be used to test how antimalarial drugs affect gametocyte production. The stage-specific determination of young vs. mature gametocytes would enable the differentiation between drugs that induce conversion versus those that select for older gametocytes that were already present prior to drug treatment.

4.2.4 Future applications for this framework: inferring infectivity from gene expression

As molecular methods have dramatically increased the sensitivity of mature gametocyte detection from thick smear microscopy-based assessments, higher rates of mature gametocyte carriage have recently observed in patient cohorts. The mere presence of mature gametocytes in the blood stream does not necessarily equate to transmissibility. Thus, more refined biomarkers of gametocyte infectivity are required to obtain a useful measurement of one's transmission potential.

Our framework could be applied to different kinds of gene expression-based inferences to achieve this goal. For example, the proportion of male and female gametocytes (both of which are required for successful transmission) could be inferred based on expression of male- and

female-specific genes. These markers could then be used in a qRT-PCR assay and could provide a way to grade infections based on infectiousness.

REFERENCES

1. WHO (2013) World Malaria Report 2012 Fact Sheet.

2. Guerra CA, Gikandi PW, Tatem AJ, Noor AM, Smith DL, et al. (2008) The limits and intensity of Plasmodium falciparum transmission: implications for malaria control and elimination worldwide. PLoS Med 5: e38.

3. Smith DL, McKenzie FE, Snow RW, Hay SI (2007) Revisiting the basic reproductive number for malaria and its implications for malaria control. PLoS Biol 5: e42.

4. Alonso PL, Brown G, Arevalo-Herrera M, Binka F, Chitnis C, et al. A research agenda to underpin malaria eradication. PLoS Med 8: e1000406.

5. Eichner M, Diebner HH, Molineaux L, Collins WE, Jeffery GM, et al. (2001) Genesis, sequestration and survival of Plasmodium falciparum gametocytes: parameter estimates from fitting a model to malariatherapy data. Trans R Soc Trop Med Hyg 95: 497-501.

6. Garcia JE, Puentes A, Patarroyo ME (2006) Developmental biology of sporozoite-host interactions in Plasmodium falciparum malaria: implications for vaccine design. Clin Microbiol Rev 19: 686-707.

7. A research agenda for malaria eradication: drugs. PLoS Med 8: e1000402.

8. Marchiafava E, Bignami A (1894) Sulle febbri estivo aumnali. E Loescher.

9. Bignami A, Bastianelli G (1890) Osservazioni sulle febbri malariche estive autunalle. La Riforma Medica 232: 1334-1335.

10. Buffet PA, Safeukui I, Deplaine G, Brousse V, Prendki V, et al. The pathogenesis of Plasmodium falciparum malaria in humans: insights from splenic physiology. Blood.

11. Seydel KB, Milner DA, Jr., Kamiza SB, Molyneux ME, Taylor TE (2006) The distribution and intensity of parasite sequestration in comatose Malawian children. J Infect Dis 194: 208-205.

12. Genrich GL, Guarner J, Paddock CD, Shieh WJ, Greer PW, et al. (2007) Fatal malaria infection in travelers: novel immunohistochemical assays for the detection of Plasmodium falciparum in tissues and implications for pathogenesis. Am J Trop Med Hyg 76: 251-259.

13. MacPherson GG, Warrell MJ, White NJ, Looareesuwan S, Warrell DA (1985) Human cerebral malaria. A quantitative ultrastructural analysis of parasitized erythrocyte sequestration. Am J Pathol 119: 385-401.

14. Pongponratn E, Turner GD, Day NP, Phu NH, Simpson JA, et al. (2003) An ultrastructural study of the brain in fatal Plasmodium falciparum malaria. Am J Trop Med Hyg 69: 345-359.

15. Nguansangiam S, Day NP, Hien TT, Mai NT, Chaisri U, et al. (2007) A quantitative ultrastructural study of renal pathology in fatal Plasmodium falciparum malaria. Trop Med Int Health 12: 1037-1050.

16. Craig A, Scherf A (2001) Molecules on the surface of the Plasmodium falciparum infected erythrocyte and their role in malaria pathogenesis and immune evasion. Mol Biochem Parasitol 115: 129-143.

17. Gardner MJ, Hall N, Fung E, White O, Berriman M, et al. (2002) Genome sequence of the human malaria parasite Plasmodium falciparum. Nature 419: 498-511.

18. Smith JD, Chitnis CE, Craig AG, Roberts DJ, Hudson-Taylor DE, et al. (1995) Switches in expression of Plasmodium falciparum var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. Cell 82: 101-110.

19. Salanti A, Dahlback M, Turner L, Nielsen MA, Barfod L, et al. (2004) Evidence for the involvement of VAR2CSA in pregnancy-associated malaria. J Exp Med 200: 1197-1203.

20. Kilejian A, Abati A, Trager W (1977) Plasmodium falciparum and Plasmodium coatneyi: immunogenicity of "knob-like protrusions" on infected erythrocyte membranes. Exp Parasitol 42: 157-164.

21. Aikawa M, Kamanura K, Shiraishi S, Matsumoto Y, Arwati H, et al. (1996) Membrane knobs of unfixed Plasmodium falciparum infected erythrocytes: new findings as revealed by atomic force microscopy and surface potential spectroscopy. Exp Parasitol 84: 339-343.

22. Crabb BS, Cooke BM, Reeder JC, Waller RF, Caruana SR, et al. (1997) Targeted gene disruption shows that knobs enable malaria-infected red cells to cytoadhere under physiological shear stress. Cell 89: 287-296.

23. Thomson JG, Robertson A (1935) The Structure and development of Plasmodium falciparum gametocytes in the internal organs and peripheral circulation. Trans R Soc Trop Med Hyg 14: 31-40.

24. Sinden RE, Canning EU, Bray RS, Smalley ME (1978) Gametocyte and gamete development in Plasmodium falciparum. Proc R Soc Lond B Biol Sci 201: 375-399.

25. Smalley ME, Abdalla S, Brown J (1981) The distribution of Plasmodium falciparum in the peripheral blood and bone marrow of Gambian children. Trans R Soc Trop Med Hyg 75: 103-105.

26. Farfour E, Charlotte F, Settegrana C, Miyara M, Buffet P The extravascular compartment of the bone marrow: a niche for Plasmodium falciparum gametocyte maturation? Malar J 11: 285.

27. Abdulsalam AH, Sabeeh N, Bain BJ Immature Plasmodium falciparum gametocytes in bone marrow. Am J Hematol 85: 943.

28. Beeson JG, Amin N, Kanjala M, Rogerson SJ (2002) Selective accumulation of mature asexual stages of Plasmodium falciparum-infected erythrocytes in the placenta. Infect Immun 70: 5412-5415.

29. Sinden RE (1982) Gametocytogenesis of Plasmodium falciparum in vitro: ultrastructural observations on the lethal action of chloroquine. Ann Trop Med Parasitol 76: 15-23.

30. Day KP, Hayward RE, Smith D, Culvenor JG (1998) CD36-dependent adhesion and knob expression of the transmission stages of Plasmodium falciparum is stage specific. Mol Biochem Parasitol 93: 167-177.

31. Silvestrini F, Tiburcio M, Bertuccini L, Alano P (2012) Differential adhesive properties of sequestered asexual and sexual stages of Plasmodium falciparum on human endothelial cells are tissue independent. PloS one 7: e31567.

32. Rogers NJ, Daramola O, Targett GA, Hall BS (1996) CD36 and intercellular adhesion molecule 1 mediate adhesion of developing Plasmodium falciparum gametocytes. Infect Immun 64: 1480-1483.

33. Tiburcio M, Silvestrini F, Bertuccini L, Sander AF, Turner L, et al. (2012) Early gametocytes of the malaria parasite Plasmodium falciparum specifically remodel the adhesive properties of infected erythrocyte surface. Cell Microbiol.

34. Silvestrini F, Lasonder E, Olivieri A, Camarda G, van Schaijk B, et al. (2010) Protein export marks the early phase of gametocytogenesis of the human malaria parasite Plasmodium falciparum. Mol Cell Proteomics 9: 1437-1448.

35. Sharp S, Lavstsen T, Fivelman QL, Saeed M, McRobert L, et al. (2006) Programmed transcription of the var gene family, but not of stevor, in Plasmodium falciparum gametocytes. Eukaryot Cell 5: 1206-1214.

36. Rogers NJ, Hall BS, Obiero J, Targett GA, Sutherland CJ (2000) A model for sequestration of the transmission stages of Plasmodium falciparum: adhesion of gametocyte-infected erythrocytes to human bone marrow cells. Infect Immun 68: 3455-3462.

37. Hanssen E, Knoechel C, Dearnley M, Dixon MW, Le Gros M, et al. (2012) Soft X-ray microscopy analysis of cell volume and hemoglobin content in erythrocytes infected with asexual and sexual stages of Plasmodium falciparum. J Struct Biol 177: 224-232.

38. Aingaran M, Zhang R, Law SK, Peng Z, Undisz A, et al. (2012) Host cell deformability is linked to transmission in the human malaria parasite Plasmodium falciparum. Cellular microbiology.

39. Tiburcio M, Niang M, Deplaine G, Perrot S, Bischoff E, et al. (2012) A switch in infected erythrocyte deformability at the maturation and blood circulation of Plasmodium falciparum transmission stages. Blood.

40. Okell LC, Drakeley CJ, Bousema T, Whitty CJ, Ghani AC (2008) Modelling the impact of artemisinin combination therapy and long-acting treatments on malaria transmission intensity. PLoS Med 5: e226; discussion e226.

41. Targett G, Drakeley C, Jawara M, von Seidlein L, Coleman R, et al. (2001) Artesunate reduces but does not prevent posttreatment transmission of Plasmodium falciparum to Anopheles gambiae. J Infect Dis 183: 1254-1259.

42. Adjalley SH, Johnston GL, Li T, Eastman RT, Ekland EH, et al. (2011) Quantitative assessment of Plasmodium falciparum sexual development reveals potent transmission-blocking activity by methylene blue. Proceedings of the National Academy of Sciences of the United States of America 108: E1214-1223.

43. Bousema T, Okell L, Shekalaghe S, Griffin JT, Omar S, et al. (2010) Revisiting the circulation time of Plasmodium falciparum gametocytes: molecular detection methods to estimate the duration of gametocyte carriage and the effect of gametocytocidal drugs. Malaria journal 9: 136.

44. Smithuis F, Kyaw MK, Phe O, Win T, Aung PP, et al. Effectiveness of five artemisinin combination regimens with or without primaquine in uncomplicated falciparum malaria: an open-label randomised trial. Lancet Infect Dis 10: 673-681.

45. Burgess RW, Bray RS (1961) The effect of a single dose of primaquine on the gametocytes, gametogony and sporogony of Laverania falciparum. Bull World Health Organ 24: 451-456.

46. Howes RE, Battle KE, Satyagraha AW, Baird JK, Hay SI G6PD deficiency: global distribution, genetic variants and primaquine therapy. Adv Parasitol 81: 133-201.

47. Karl S, David M, Moore L, Grimberg BT, Michon P, et al. (2008) Enhanced detection of gametocytes by magnetic deposition microscopy predicts higher potential for Plasmodium falciparum transmission. Malar J 7: 66.

48. Niederwieser I, Felger I, Beck HP (2000) Plasmodium falciparum: expression of gametocyte-specific genes in monolayer cultures and malaria-positive blood samples. Experimental parasitology 95: 163-169.

49. Schneider P, Schoone G, Schallig H, Verhage D, Telgt D, et al. (2004) Quantification of Plasmodium falciparum gametocytes in differential stages of development by quantitative nucleic acid sequence-based amplification. Mol Biochem Parasitol 137: 35-41.

50. Buates S, Bantuchai S, Sattabongkot J, Han ET, Tsuboi T, et al. Development of a reverse transcription-loop-mediated isothermal amplification (RT-LAMP) for clinical detection of Plasmodium falciparum gametocytes. Parasitol Int 59: 414-420.

51. Schneider P, Bousema T, Omar S, Gouagna L, Sawa P, et al. (2006) (Sub)microscopic Plasmodium falciparum gametocytaemia in Kenyan children after treatment with sulphadoxine-pyrimethamine monotherapy or in combination with artesunate. Int J Parasitol 36: 403-408.

52. Ouedraogo AL, Schneider P, de Kruijf M, Nebie I, Verhave JP, et al. (2007) Age-dependent distribution of Plasmodium falciparum gametocytes quantified by Pfs25 real-time QT-NASBA in a cross-sectional study in Burkina Faso. Am J Trop Med Hyg 76: 626-630.

53. Shekalaghe SA, Bousema JT, Kunei KK, Lushino P, Masokoto A, et al. (2007) Submicroscopic Plasmodium falciparum gametocyte carriage is common in an area of low and seasonal transmission in Tanzania. Trop Med Int Health 12: 547-553.

54. Bousema JT, Schneider P, Gouagna LC, Drakeley CJ, Tostmann A, et al. (2006) Moderate effect of artemisinin-based combination therapy on transmission of Plasmodium falciparum. The Journal of infectious diseases 193: 1151-1159.

55. Mens PF, Sawa P, van Amsterdam SM, Versteeg I, Omar SA, et al. (2008) A randomized trial to monitor the efficacy and effectiveness by QT-NASBA of artemether-lumefantrine versus dihydroartemisinin-piperaquine for treatment and transmission control of uncomplicated Plasmodium falciparum malaria in western Kenya. Malaria journal 7: 237.

56. Shekalaghe S, Drakeley C, Gosling R, Ndaro A, van Meegeren M, et al. (2007) Primaquine clears submicroscopic Plasmodium falciparum gametocytes that persist after treatment with sulphadoxine-pyrimethamine and artesunate. PloS one 2: e1023.

57. Niederwieser I, Felger I, Beck HP (2001) Limited polymorphism in Plasmodium falciparum sexual-stage antigens. The American journal of tropical medicine and hygiene 64: 9-11.

58. Bousema JT, Schneider P, Gouagna LC, Drakeley CJ, Tostmann A, et al. (2006) Moderate effect of artemisinin-based combination therapy on transmission of Plasmodium falciparum. J Infect Dis 193: 1151-1159.

59. Ouedraogo AL, Bousema T, de Vlas SJ, Cuzin-Ouattara N, Verhave JP, et al. The plasticity of Plasmodium falciparum gametocytaemia in relation to age in Burkina Faso. Malar J 9: 281.

60. Drakeley CJ, Akim NI, Sauerwein RW, Greenwood BM, Targett GA (2000) Estimates of the infectious reservoir of Plasmodium falciparum malaria in The Gambia and in Tanzania. Trans R Soc Trop Med Hyg 94: 472-476.

61. Alano P (2007) Plasmodium falciparum gametocytes: still many secrets of a hidden life. Mol Microbiol 66: 291-302.

62. Price R, Nosten F, Simpson JA, Luxemburger C, Phaipun L, et al. (1999) Risk factors for gametocyte carriage in uncomplicated falciparum malaria. Am J Trop Med Hyg 60: 1019-1023.

63. Bousema T, Drakeley C (2011) Epidemiology and infectivity of Plasmodium falciparum and Plasmodium vivax gametocytes in relation to malaria control and elimination. Clinical microbiology reviews 24: 377-410.

64. Bousema T, Drakeley C Epidemiology and infectivity of Plasmodium falciparum and Plasmodium vivax gametocytes in relation to malaria control and elimination. Clin Microbiol Rev 24: 377-410.

65. Metselaar D (1960) Relative increase in the prevalence of Plasmodium falciparum some years after the beginning of a house-spraying campaign in Netherlands New Guinea. Trans R Soc Trop Med Hyg 54: 523-528.

66. Molineaux LaG, G. (1980) The Garki Project. Research on the Epidemiology and Control of Malaria in the Sudan Savanna of West Africa. World Health Organization. Geneva, Switzerland.

67. Miller LH, Baruch DI, Marsh K, Doumbo OK (2002) The pathogenic basis of malaria. Nature 415: 673-679.

68. Hiller NL, Bhattacharjee S, van Ooij C, Liolios K, Harrison T, et al. (2004) A host-targeting signal in virulence proteins reveals a secretome in malarial infection. Science 306: 1934-1937.

69. Maier AG, Rug M, O'Neill MT, Brown M, Chakravorty S, et al. (2008) Exported proteins required for virulence and rigidity of Plasmodium falciparum-infected human erythrocytes. Cell 134: 48-61.

70. Marti M, Good RT, Rug M, Knuepfer E, Cowman AF (2004) Targeting malaria virulence and remodeling proteins to the host erythrocyte. Science 306: 1930-1933.

71. Rogers NJ, Targett GA, Hall BS (1996) Plasmodium falciparum gametocyte adhesion to C32 cells via CD36 is inhibited by antibodies to modified band 3. Infect Immun 64: 4261-4268.

72. Dixon MW, Dearnley MK, Hanssen E, Gilberger T, Tilley L Shape-shifting gametocytes: how and why does P. falciparum go banana-shaped? Trends Parasitol 28: 471-478.

73. Taylor TE, Fu WJ, Carr RA, Whitten RO, Mueller JS, et al. (2004) Differentiating the pathologies of cerebral malaria by postmortem parasite counts. Nat Med 10: 143-145.

74. Fivelman QL, McRobert L, Sharp S, Taylor CJ, Saeed M, et al. (2007) Improved synchronous production of Plasmodium falciparum gametocytes in vitro. Mol Biochem Parasitol 154: 119-123.

75. van der Loos CM (2008) Multiple immunoenzyme staining: methods and visualizations for the observation with spectral imaging. J Histochem Cytochem 56: 313-328.

76. Eksi S, Stump A, Fanning SL, Shenouda MI, Fujioka H, et al. (2002) Targeting and sequestration of truncated Pfs230 in an intraerythrocytic compartment during Plasmodium falciparum gametocytogenesis. Mol Microbiol 44: 1507-1516.

77. Eksi S, Czesny B, van Gemert GJ, Sauerwein RW, Eling W, et al. (2006) Malaria transmission-blocking antigen, Pfs230, mediates human red blood cell binding to exflagellating male parasites and oocyst production. Mol Microbiol 61: 991-998.

78. Eksi S, Williamson KC (2011) Protein targeting to the parasitophorous vacuole membrane of Plasmodium falciparum. Eukaryotic cell 10: 744-752.

79. Smith JD, Kyes S, Craig AG, Fagan T, Hudson-Taylor D, et al. (1998) Analysis of adhesive domains from the A4VAR Plasmodium falciparum erythrocyte membrane protein-1 identifies a CD36 binding domain. Mol Biochem Parasitol 97: 133-148.

80. Daily JP, Scanfeld D, Pochet N, Le Roch K, Plouffe D, et al. (2007) Distinct physiological states of Plasmodium falciparum in malaria-infected patients. Nature 450: 1091-1095.

81. Prommano O, Chaisri U, Turner GD, Wilairatana P, Ferguson DJ, et al. (2005) A quantitative ultrastructural study of the liver and the spleen in fatal falciparum malaria. Southeast Asian J Trop Med Public Health 36: 1359-1370.

82. Young JA, Fivelman QL, Blair PL, de la Vega P, Le Roch KG, et al. (2005) The Plasmodium falciparum sexual development transcriptome: a microarray analysis using ontology-based pattern identification. Mol Biochem Parasitol 143: 67-79.

83. Pradel G (2007) Proteins of the malaria parasite sexual stages: expression, function and potential for transmission blocking strategies. Parasitology 134: 1911-1929.

84. Wickramasinghe SN, Phillips RE, Looareesuwan S, Warrell DA, Hughes M (1987) The bone marrow in human cerebral malaria: parasite sequestration within sinusoids. British journal of haematology 66: 295-306.

85. Healer J, Graszynski A, Riley E (1999) Phagocytosis does not play a major role in naturally acquired transmission-blocking immunity to Plasmodium falciparum malaria. Infect Immun 67: 2334-2339.

86. Chasis JA, Mohandas N (2008) Erythroblastic islands: niches for erythropoiesis. Blood 112: 470-478.

87. Trager W, Gill GS, Lawrence C, Nagel RL (1999) Plasmodium falciparum: enhanced gametocyte formation in vitro in reticulocyte-rich blood. Exp Parasitol 91: 115-118.

88. Gautret P, Miltgen F, Gantier JC, Chabaud AG, Landau I (1996) Enhanced gametocyte formation by Plasmodium chabaudi in immature erythrocytes: pattern of production, sequestration, and infectivity to mosquitoes. J Parasitol 82: 900-906.

89. Talman AM, Blagborough AM, Sinden RE (2010) A Plasmodium falciparum strain expressing GFP throughout the parasite's life-cycle. PloS one 5: e9156.

90. Bei AK, Brugnara C, Duraisingh MT (2010) In vitro genetic analysis of an erythrocyte determinant of malaria infection. J Infect Dis 202: 1722-1727.

91. Wickramasinghe SN, Phillips RE, Looareesuwan S, Warrell DA, Hughes M (1987) The bone marrow in human cerebral malaria: parasite sequestration within sinusoids. Br J Haematol 66: 295-306.

92. Dearnley MK, Yeoman JA, Hanssen E, Kenny S, Turnbull L, et al. (2012) Origin, composition, organization and function of the inner membrane complex of Plasmodium falciparum gametocytes. Journal of cell science.

93. Saeed M, Roeffen W, Alexander N, Drakeley CJ, Targett GA, et al. (2008) Plasmodium falciparum antigens on the surface of the gametocyte-infected erythrocyte. PloS one 3: e2280.

94. Marsee DK, Pinkus GS, Yu H CD71 (transferrin receptor): an effective marker for erythroid precursors in bone marrow biopsy specimens. Am J Clin Pathol 134: 429-435.

95. Trager W, Gill GS (1992) Enhanced gametocyte formation in young erythrocytes by Plasmodium falciparum in vitro. J Protozool 39: 429-432.

96. Bozdech Z, Llinas M, Pulliam BL, Wong ED, Zhu J, et al. (2003) The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. PLoS Biol 1: E5.

97. Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, et al. (2003) Discovery of gene function by expression profiling of the malaria parasite life cycle. Science 301: 1503-1508.

98. Eksi S, Haile Y, Furuya T, Ma L, Su X, et al. (2005) Identification of a subtelomeric gene family expressed during the asexual-sexual stage transition in Plasmodium falciparum. Mol Biochem Parasitol 143: 90-99.

99. Silvestrini F, Bozdech Z, Lanfrancotti A, Di Giulio E, Bultrini E, et al. (2005) Genome-wide identification of genes upregulated at the onset of gametocytogenesis in Plasmodium falciparum. Mol Biochem Parasitol 143: 100-110.

100. Clarke J, Seo P, Clarke B Statistical expression deconvolution from mixed tissue samples. Bioinformatics 26: 1043-1049.

101. Stuart RO, Wachsman W, Berry CC, Wang-Rodriguez J, Wasserman L, et al. (2004) In silico dissection of cell-type-associated patterns of gene expression in prostate cancer. Proc Natl Acad Sci U S A 101: 615-620.

102. Bar-Joseph Z, Farkash S, Gifford DK, Simon I, Rosenfeld R (2004) Deconvolving cell cycle expression data with complementary information. Bioinformatics 20 Suppl 1: i23-30.

103. Taylor TE (2009) Caring for children with cerebral malaria: insights gleaned from 20 years on a research ward in Malawi. Trans R Soc Trop Med Hyg 103 Suppl 1: S6-10.

104. Milner DA, Jr., Pochet N, Krupka M, Williams C, Seydel K, et al. (2012) Transcriptional Profiling of Plasmodium falciparum Parasites from Patients with Severe Malaria Identifies Distinct Low vs. High Parasitemic Clusters. PloS one 7: e40739.

105. Buchholz K, Burke TA, Williamson KC, Wiegand RC, Wirth DF, et al. (2011) A high-throughput screen targeting malaria transmission stages opens new avenues for drug development. J Infect Dis 203: 1445-1453.

106. Alano P, Roca L, Smith D, Read D, Carter R, et al. (1995) Plasmodium falciparum: parasites defective in early stages of gametocytogenesis. Exp Parasitol 81: 227-235.

107. Trager W, Jensen JB (1976) Human malaria parasites in continuous culture. Science 193: 673-675.

108. Lambros C, Vanderberg JP (1979) Synchronization of Plasmodium falciparum erythrocytic stages in culture. J Parasitol 65: 418-420.

109. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, et al. (2006) GenePattern 2.0. Nat Genet 38: 500-501.

110. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics 25: 25-29.

111. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic acids research 38: D355-360.

112. Merrick CJ, Huttenhower C, Buckee C, Amambua-Ngwa A, Gomez-Escobar N, et al. (2012) Epigenetic Dysregulation of Virulence Gene Expression in Severe Plasmodium falciparum Malaria. The Journal of infectious diseases 205: 1593-1600.

113. Van Tyne D, Park DJ, Schaffner SF, Neafsey DE, Angelino E, et al. (2011) Identification and Functional Validation of the Novel Antimalarial Resistance Locus PF10_0355 in Plasmodium falciparum. PLoS Genet 7: e1001383.

114. Neafsey DE, Schaffner SF, Volkman SK, Park D, Montgomery P, et al. (2008) Genome-wide SNP genotyping highlights the role of natural selection in Plasmodium falciparum population divergence. Genome Biol 9: R171.

115. Walliker D, Quakyi IA, Wellems TE, McCutchan TF, Szarfman A, et al. (1987) Genetic analysis of the human malaria parasite Plasmodium falciparum. Science 236: 1661-1666.

116. Blisnick T, Morales Betoulle ME, Barale JC, Uzureau P, Berry L, et al. (2000) Pfsbp1, a Maurer's cleft Plasmodium falciparum protein, is associated with the erythrocyte skeleton. Mol Biochem Parasitol 111: 107-121.

117. Sargeant TJ, Marti M, Caler E, Carlton JM, Simpson K, et al. (2006) Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. Genome Biol 7: R12.

118. van Wolfswinkel ME, de Mendonca Melo M, Vliegenthart-Jongbloed K, Koelewijn R, van Hellemond JJ, et al. The prognostic value of schizontaemia in imported Plasmodium falciparum malaria. Malar J 11: 301.

119. Booker ML, Bastos CM, Kramer ML, Barker RH, Jr., Skerlj R, et al. Novel inhibitors of Plasmodium falciparum dihydroorotate dehydrogenase with anti-malarial activity in the mouse model. J Biol Chem 285: 33054-33064.

120. Okamoto N, Spurck TP, Goodman CD, McFadden GI (2009) Apicoplast and mitochondrion in gametocytogenesis of Plasmodium falciparum. Eukaryotic cell 8: 128-132.

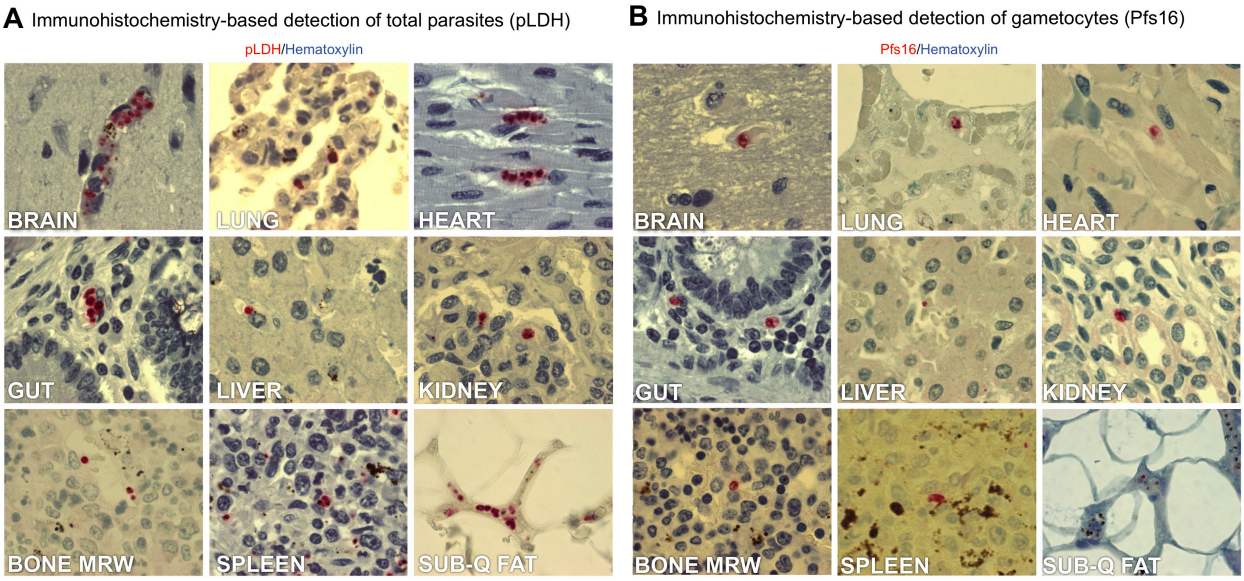
121. Ouedraogo AL, Bousema T, Schneider P, de Vlas SJ, Ilboudo-Sanogo E, et al. (2009) Substantial contribution of submicroscopical Plasmodium falciparum gametocyte carriage to the infectious reservoir in an area of seasonal transmission. PLoS One 4: e8410.

122. Drakeley CJ, Secka I, Correa S, Greenwood BM, Targett GA (1999) Host haematological factors influencing the transmission of Plasmodium falciparum gametocytes to Anopheles gambiae s.s. mosquitoes. Trop Med Int Health 4: 131-138.

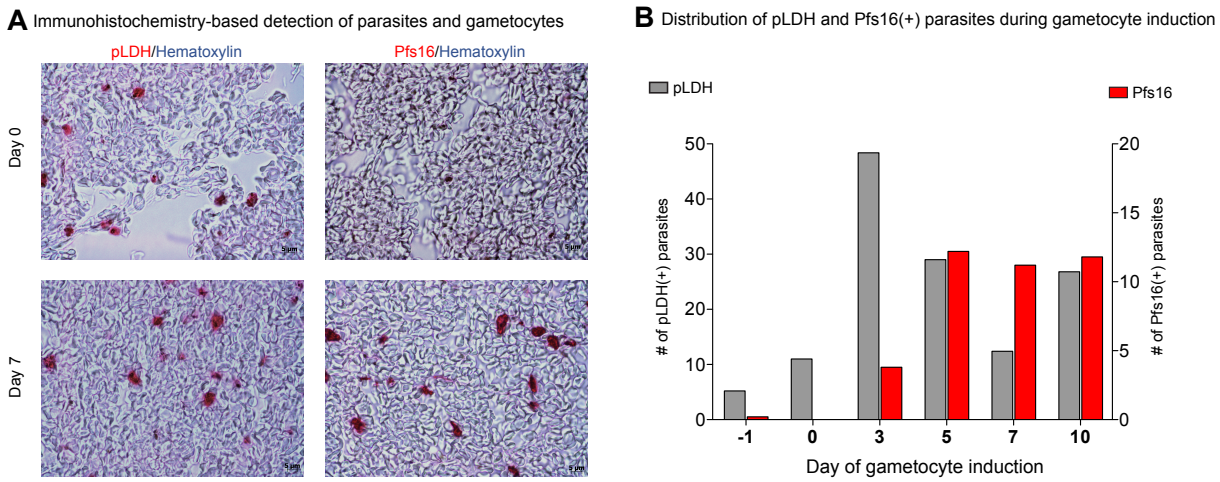
123. Hanssen E, Knoechel C, Dearnley M, Dixon MW, Le Gros M, et al. (2011) Soft X-ray microscopy analysis of cell volume and hemoglobin content in erythrocytes infected with asexual and sexual stages of Plasmodium falciparum. Journal of structural biology.

124. Bachmann A, Esser C, Petter M, Predehl S, von Kalckreuth V, et al. (2009) Absence of erythrocyte sequestration and lack of multicopy gene family expression in Plasmodium falciparum from a splenectomized malaria patient. PLoS ONE 4: e7459.

APPENDIX A



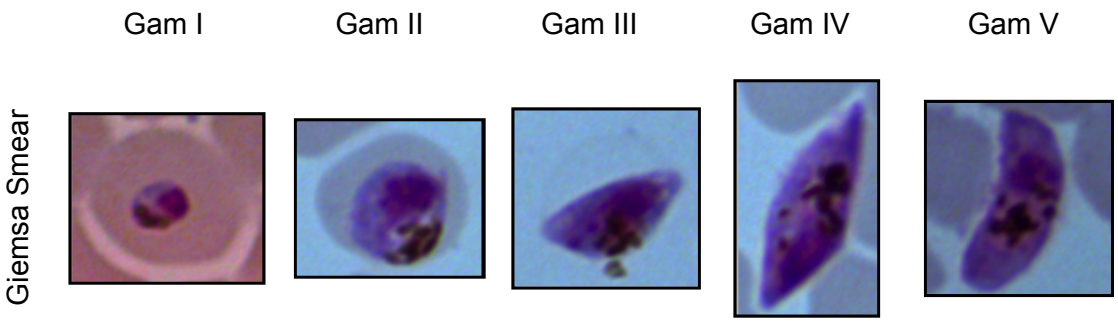
Appendix Figure A1. Representative pLDH and Pfs16 labeling across tissues. (A). pLDH. pLDH-labeled parasites were identified in all the 9 tissue types analyzed. In a subset of those tissue (i.e., brain, gut and heart) the classically described pathology of packed vessels was observed. (**B). Pfs16.** Pfs16-labeled parasites were observed in multiple tissues including those previously reported with gametocytes (i.e., bone marrow and spleen).



Appendix Figure A2. Stage Specificity of antibodies by IHC. Gametocytes were induced according an established *in vitro* gametocye induction protocol [74], with Day 0 being the first day of gametocyte development. Infected RBCs from the time course were mixed with surgical tissues fixed in formalin and embedded in paraffin (FFPE) to be used for optimization and positive controls in subsequent experiments. (**A). Images of parasites labeled with either pLDH or Pfs16.** Representative images of IHC with pLDH and Pfs16 on *in vitro* control tissue blocks are shown from Day 0 and Day 7. (**B). Quantification of antibody labeling across gametocyte development.** The average number of Fast Red-positives observed in 5 high power fields (HPFs) is shown. As expected, Pfs16 labeling is present at Day 7 of gametocyte development but not at Day 0 (where parasites are still at ring stage). Of note, given the 1:5000 antibody titration used, we did not observe Pfs16 labeling in Day -1 (committed schizonts) or Day 0 (gametocyte ring). We began seeing Pfs16 staining at Day 2 of small oat-shaped stage I gametocytes.

Appendix Figure A3. Morphology of gametocyte stages by immunohistochemistry. Shown are *in vitro* cultured infected RBCs from the gametocyte induction protocol used to generate RNA and IHC control samples. Stage distribution in each sample was determined using methanol fixed smears followed by Wright's Giemsa stain. Images of stages I-V, as determined by morphology, are shown from the Giemsa stain. Based on known distributions of stages, images of gametocytes in each stage were imaged from the FFPE control block (containing in vitro cultured gametocytes from the same daily time points as the Giemsa stained slides). These were taken following IHC using mouse Pfs16 antibodies with AP-Fast Red. Images were then processed in ImageJ using the particle analysis function. Measurements of area in square micron and length in micron are shown for the range of gametocytes imaged for that stage. For each stage, we had at least five measurements. Shown is the median (and range).

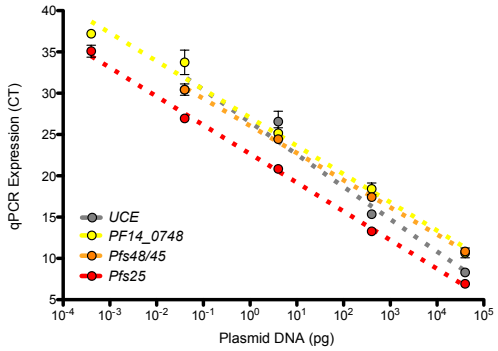
Detection of gametocytes by Immunohistochemistry (Pfs16)



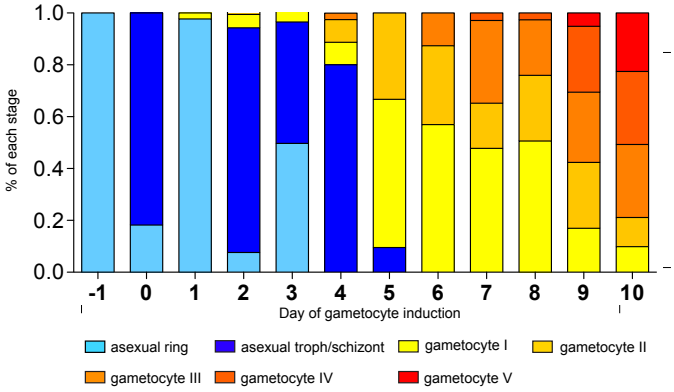
	Gam I	Gam II	Gam III	Gam IV	Gam V
area (sq micron)	6.7 (2.4 - 13.8)	17.8 (7.9-24.0)	27.9 (14.4-34.7)	33.5 (23.0-37.6)	36.3 (32.6-38.2)
length (micron)	4.0 (2.1 - 5.2)	5.8 (4.4-7.5)	8.1 (6.9-9.6)	10.6 (9.7-12.4)	10.5 (10.1-13.2)

Appendix Figure A4. Development of qRT-PCR assay. (A). Primer efficiency. Each amplicon was inserted into a pGEM-T easy plasmid and curves generated using serial dilutions of plasmid DNA (see Appendix Table A1 for primers). **(B). Gametocyte time course and staging.** Gametocytes were induced according to Fivelman et al [74] and samples collected every day for 10 days. Parasites were labeled with nuclear dye DAPI and gametocytes were stained with antibodies against the gametocyte marker Pfs16. Gametocytes were identified based on Pfs16-labeling and categorized into the five known stages I-V, according to Sinden et al [29]. **(C). Representative images from the time course experiments.** Shown are early, mid stage and late gametocytes, labeled with Pfs16 antibody and DAPI. **(D). Stage-specific expression of markers.** While UCE shows relatively constitutive expression, the gametocyte markers increase upon induction according to their known expression peaks: *PF14_0748* peaks at days 3 and 4, *Pfs48/45* peaks at days 7-9 and *Pfs25* peaks at days 9 and 10.

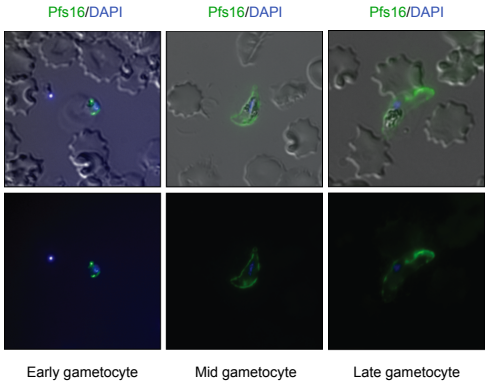
A Primer efficiency curves for stage-specific markers



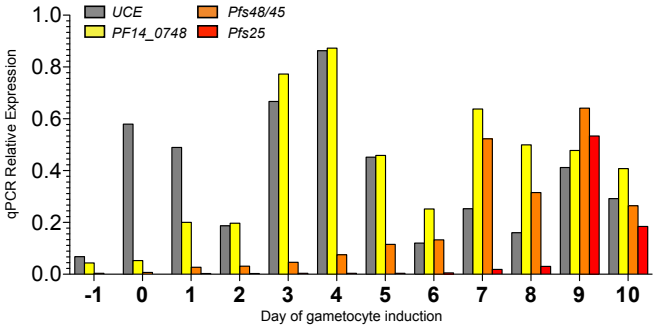
B Microscopy-based distribution of asexual and sexual stages



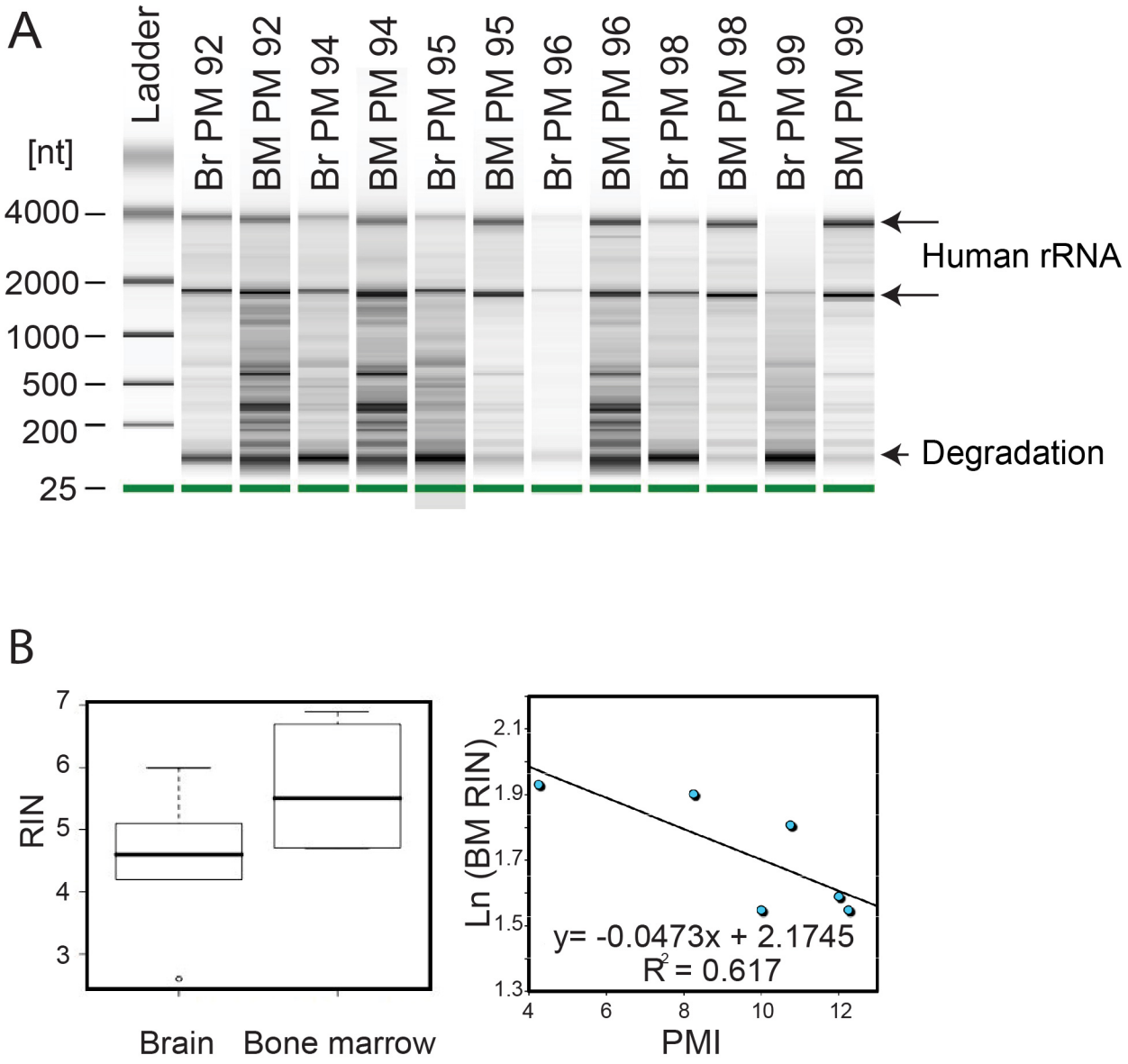
C Immunofluorescence for staging gametocytes



D Stage-specific expression of early, mid and late gametocyte genes



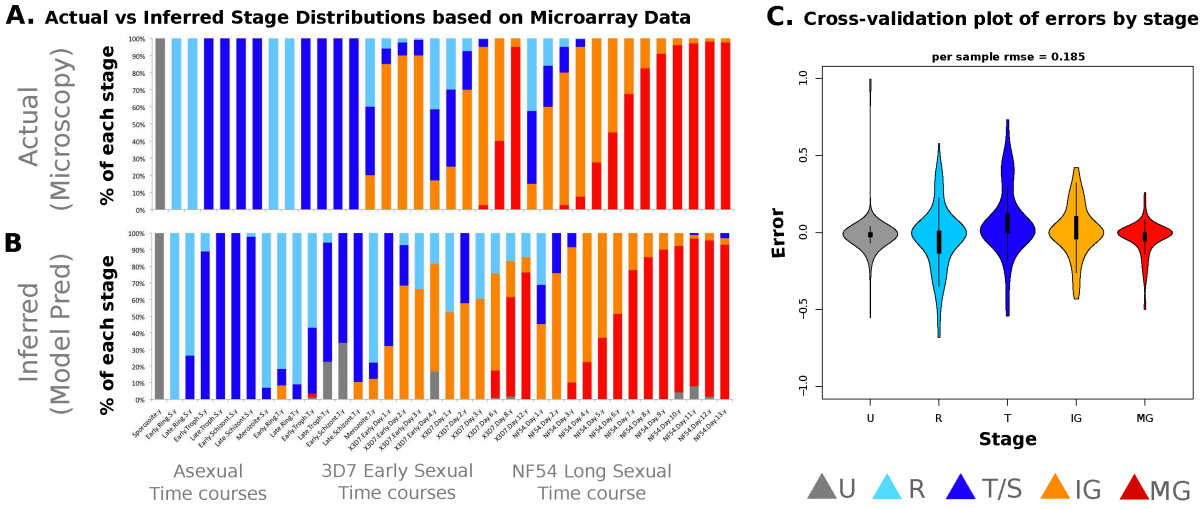
Appendix Figure A5. RNA quality from brain and bone marrow tissue samples. (A). RNA size distribution per sample. Purified and DNAse digested RNA from 6 brain (BR) and 6 bone marrow samples (BM) was run on a Bioanalyzer to determine the length distribution of RNA fragments. The most prominent bands are human 28s and 18s rRNA bands (at app 4000 and 2000nt, respectively, see arrows). (**B). RIN and PMI.** Based on the average RNA integrity number (RIN, a measure for the average RNA fragment length), the RNA in these tissues is considered partially degraded but sufficient for the small fragments (100 - 200 nt) amplified by the qRT-PCR assay. A positive correlation is observed between the RIN and post mortem intervals (PMI, the time between death and autopsy), suggesting that RNA degrades more with increased PMI. In addition RNA from brain samples appears generally more degraded than from bone marrow samples**.**


A
Ladder
Br PM 92
BM PM 92
Br PM 94
BM PM 94
Br PM 95
BM PM 95
Br PM 96
BM PM 96
Br PM 98
BM PM 98
Br PM 99
BM PM 99
[nt]
4000
2000
1000
500
200
25
Human rRNA
Degradation
B
RIN
7
6
5
4
3
Brain
Bone marrow
Ln (BM RIN)
2.1
1.9
1.7
1.5
1.3
y= -0.0473x + 2.1745
R2 = 0.617
4
6
8
10
12
PMI

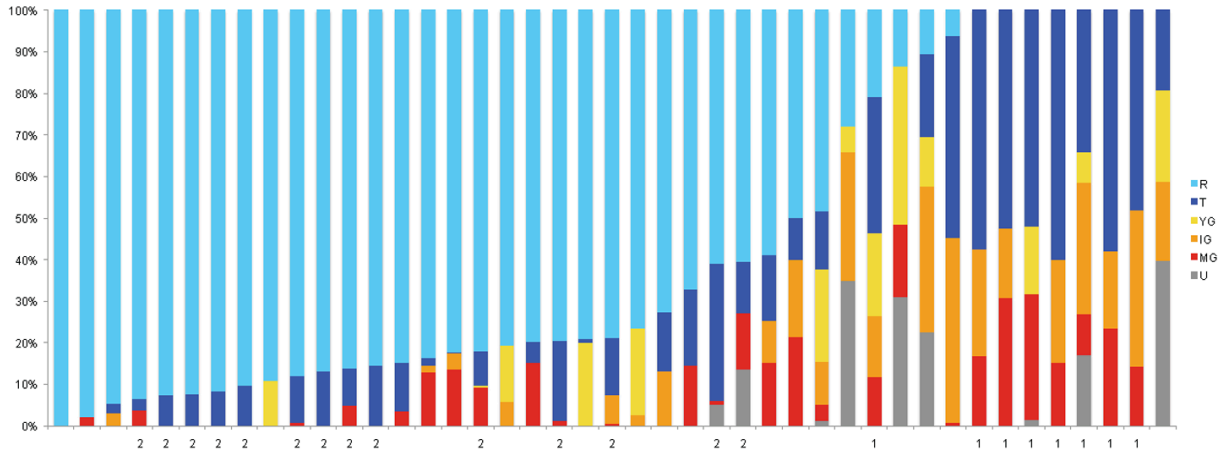
Appendix Table A1. Primers used in this study for the qRT-PCR assay.

Name	Accession	Stage-Specificity	Primer Efficiency	Primer sequence (listed 5' to 3')
Plasmodium exported protein (PHISTa)	*PF14_0748*	early - mid gametocyte	97.6%	ATTCAAGGGTAGTTCCTAGAGCAGTGTGG AGCACTCGTAATTCTAACACTGGG
6-cysteine protein (Pfs48/45)	*PF13_0247*	mid - late gametocyte	101.4%	GTAAGCCTAGCTCTTTGA ATAGTGA GACCTACGTTCACGCATATCTGGCT
25 kDa ookinete surface antigen precursor	*PF10_0303*	late gametocyte	93.1%	GGA AATCCCGTTTCATACGCTTGT TCTTGTACATTGGGAACTTTGCCT
Ubiquitin-conjugating enzyme	*PF08_0085*	all stages	100.5%	GGTGTTAGTGGCTCACCAATAGGA GTACCACCTTCCCATGGAGTATCA

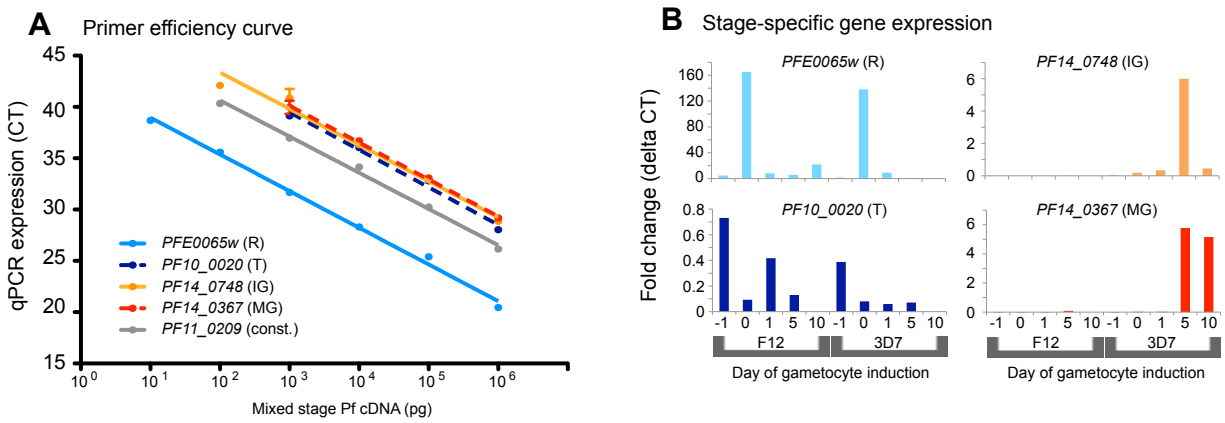
APPENDIX B



Appendix Figure B1. Performance of the 5 marker model on published microarray data sets. (A). Actual and **(B)** inferred stage distributions across five microarray time courses (two asexual and three sexual) with reference stage distributions determined by microscopy. Five markers were used to make these predictions (Table1). **(C).** Bootstrap cross-validation of error rates expected per-stage in model inferences. Violin plots show expected density, with internal boxplots detailing the 25th -75 th percentiles and 1.5x fences.



Appendix Figure B2. Application of the model on microarray data from previously analyzed patient samples. Re-analysis of an uncomplicated patient cohort collected in Senegal between 2004 - 2006, and published in [80]. Proportions of total parasitemia as inferred by modeling of parasite transcriptional activity in patients are shown here, as measured by hybridization on Affymetrix chip. As expected, most uncomplicated patients are dominated by asexual ring stages, with minorities in both groups transitioning to trophozoites. Disproportionately large fractions of gametocytes and uncharacterized stages enriched in the Cluster 1 patients.



Appendix Figure B3. qRT-PCR Assay Optimization. (A). Efficiency of qRT-PCR reactions using 10-fold dilutions of mixed parasite cDNA. 4 to 6 dilutions were assessed for each primer, and efficiencies were in the acceptable range for all 5 primers (87-92%). R^2-values were all greater than 0.96. Technical variation between replicates was very low: the average standard deviation between technical replicates was 0.243 and ranged between 0.01 and 1.543. **(B).** Stage-specificity of qRT-PCR markers. Using two clones of 3D7, F12 (gametocyte deficient) and wild type (gametocyte producer), we performed *in vitro* gametocyte inductions and collected samples at days -1, 0, 1, 5, 10 according to the Fivelman et al protocol [74] and collected RNA. Results, displayed as relative expression normalized to constitutively expressed marker *PF11_0209*, confirm stage-specificity of markers. Giemsa images of representative parasite stages observed at each time point of the wild type 3D7 time course are shown above the graphs

[TableB1.xlsx]

Appendix Table B1: Annotated Gene List and Metadata. Annotations of the 5160 genes in the *P. falciparum* transcriptome used in the analysis, including the frequency of selection in our subsampling and backward selection steps, presence of an intron, contribution of expression to stage, determination of stage specificity, product description and population genetic parameters of total SNP counts, diversity and divergence.

[TableB2.xlsx]

Appendix Table B2: Complete GSEA Results per Stage. Results for each stage in our microarray model, wherein the per gene z-scored contributions of expression to that stage were ranked and were characterized for enrichment in functional pathways.

[TableB3.xlsx]

Appendix Table B3: GSEA Gene Sets. Gold Standard Catalog of GO and Kegg pathways obtained from individual GO slims from PlasmoDB and the GO ontology integrated into the GO hierarchical structure.

Appendix Table B4. Sequences for primers and probe used in qRT-PCR assay.

PFE0065w **(ring)**	
Reverse:	TTGCTAGGTAATATCCTTTTCTTTTTCC
Forward:	GCAAAACAAGCCGTACATGTTG
Probe:	6FAM - TTG TTC ATC AAC TTT TAC AAC TT - MGBNFQ
PF10_0020 **(troph/schizont)**	
Reverse:	GACGTTTGATTTGTTTCCTGTTTTATC
Forward:	GGAATGATTTATTTGTTAATTAAAGATGTTG
Probe:	6FAM - ACG AGG AAA TTA GCT GAA GC - MGBNFQ
PF14_0748 **(early-mid gametocytes)**	
Reverse:	TTGGCCACACTGCTCTAGGA
Forward:	CTTATGTGCTGAATTTTGTGTTATGGT
Probe:	VIC - CAC ATA ATG AAT TCA AGG GTA G - MGBNFQ
PF14_0367 **(mid-late gametocytes)**	
Reverse:	TCCCTGTGTTTTTGCTCATCTTC
Forward:	GTTACATTTCGACCCAGCATAAATT
Probe:	VIC - CAG TGC ATA TTG TTG CCT GT - MGBNFQ
PF11_0209 **(all stages)**	
Reverse:	CATAATGCTACTAACTACTAATATGCAAAAATATACC
Forward:	CGCTAGAATTACATGGAGACAAATCA
Probe:	VIC - AAA AGG TCA AGC CTT CAT T - MGBNFQ

Appendix Table B5. Additional qRT-PCR Assay Optimization Data.

Gene	Stage-Specificity	Pf cDNA (mixed stage)		Pf DNA		Human cDNA		Human DNA		H2O	
All reactions run to 45 cycles.		Probe	SYBR	Probe	SYBR	Probe	SYBR	Probe	SYBR	Probe	SYBR
PFE0065w	ring	20.173	15.849	-	39.713	-	-	-	42.893	-	-
PF10_0020	trophozoite/ schizont	30.502	25.823	-	44.637	-	-	-	-	-	-
PF14_0748	early - mid gametocyte	32.287	27.037	-	40.020	-	-	-	-	-	-
PF14_0367	mid - late gametocyte	33.313	29.075	-	40.950	-	-	-	-	-	-
PF11_0209	all stages	27.530	22.576	-	-	-	-	-	-	-	-

Appendix Table B5. Primers were specifically designed to cross exon-exon junctions, so as to reduce genomic DNA amplification, and were checked for homology against *Plasmodium* or human homologous sequences using PlasmoDB and NCBI Blast in order to eliminate the chances of non-specific amplification. Using our primer set with sequence-specific probes showed no cross-reactivity with genomic DNA or human templates. Our primer sets also greatly reduced the amount of genomic DNA amplification even using SYBR (CT>39 as compared with DNA-amplifying control marker at CT = 25), yet it was not zero.